Does It Pay to Follow Anomalies Research?
Machine Learning Approach with International Evidence

Ondrej Tobek* & Martin Hronec†

June 2019

ABSTRACT

We study out-of-sample returns on 153 anomalies in equities documented in academic literature. We show that machine learning techniques that aggregate all the anomalies into one mispricing signal are four times more profitable than a strategy based on individual anomalies and survive on a liquid universe of stocks. We next study value of international evidence for selection of quantitative strategies that outperform out-of-sample. Past performance of quantitative strategies in the regions other than the US does not help to pick out-of-sample winning strategies in the US. Past evidence from the US, however, captures most of the predictability outside the US.

JEL classification: G11, G12, and G15.

*University of Cambridge, Department of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK; Email: ot234@cam.ac.uk

†Charles University, Institute of Economic Studies, Opletalova 26, 110 00, Prague 1; Email: martin.hronec@fsv.cuni.cz. This work has been supported by Charles University Research Centre program No. UNCE/HUM/035. This research was also supported by the Grant Agency of Charles University (grant no. 1270218).

Low interest rates environment after the Financial Crisis of 2008 has caused a surge in search for alternative ways of how to earn steady returns that are uncorrelated with the stocks market. One response of the financial industry was an explosion in a number of "smart beta" funds that provide exposure to various risk factors, which have been historically connected to risk premia. This larger interest should, however, in turn lead to their lower profitability. McLean and Pontiff (2016) document the decrease of 58% in post-publication returns relative to the in-sample returns of anomalies. Jacobs and Müller (2017b) however show that the United States is the only country with a reliable post-publication decline in returns of anomalies, emphasizing the importance of international evidence in asset pricing. Apart from the lack of the post-publication decline in the international setting, Jacobs and Müller (2017a) find that combining anomalies into one mispricing signal using least squares leads to superior out-of-sample risk-adjusted returns relative to focusing on individual anomalies. The benefit of combining individual anomalies through predictive regressions is further emphasized by Gu, Kelly, and Xiu (2018) who conclude that sophisticated machine learning methods offer higher out-of-sample predictability in the US compared to the traditional methods in Jacobs and Müller (2017a). This study extends the use of machine learning methods to international sample and finds internationally unprecedented out-of-sample profitability using anomalies as predictors in machine-learning-based predictive regressions.

In order to benchmark machine learning based strategy (mispricing strategy hereafter) we look at out-of-sample profitability of a portfolio-level strategy that invests in the individual published anomalies (portfolio-mixing strategy hereafter). Having all the constructed anomalies at our disposal, we examine degree of predictability of future profitability of the individual anomalies based on their past profitability in various regions. We also study the value of international evidence for the prediction of out-of-sample stock returns in the mispricing strategy. Next, we examine limits to arbitrage associated with our strategies. We are the first to extensively estimate transaction costs associated with strategies leveraging predictive power of anomalies internationally and document that strategies remain profitable even after accounting for the transaction costs as well as short-selling constraints. Since we only include anomalies as predictors after their publication we also examine the marginal value of the new anomalies for the out-of-sample predictions after accounting for the already published anomalies and show that it remains positive over time, confirming added value of recent anomalies literature.

153 published anomalies are studied in the US, Japan, Europe, and Asia Pacific. The anomalies in this study describe characteristics related to individual stocks that can predict their future returns. No distinction is being made between characteristics that are related to risk premia and characteristics that are related to mispricing due to frictions or other market imperfections. The studied anomalies are, for example, accruals of Sloan (1996), earnings over price of Basu (1977),

composite equity issuance of Daniel and Titman (2006), and R&D over Market Equity of Chan, Lakonishok, and Sougiannis (2001). The focus in this study is restricted to a liquid universe of stocks. The liquid stocks are defined as the largest stocks with capitalization in the top 90% of the overall market's capitalization and dollar trading volume over the previous year in the top 90% of the overall market's volume in the individual regions. Only about 500 most liquid stocks pass the criteria in 2010s in a given month in the US. Excluding small-capitalization stocks leads to results more relevant to investors and limits effect of microstructure noise.[1]

The portfolio-mixing strategy describing average return on the individual anomalies is first considered. The portfolio-mixing strategy equally invests in portfolios created based on individual anomalies that are significant in the US at 5% level.[2] Hou, Xue, and Zhang (2017) show that many of the published anomalies disappear on liquid universe of stocks. Our stock universe is far more liquid relative to Hou et al. (2017). The focus on significant anomalies in the strategy therefore guarantees that the conclusions are not driven by inclusion of these irrelevant strategies, as would be the case for the simplest strategy taking into account all the published anomalies in Hou et al. (2017). The weighting in the strategy is the simplest possible and the strategy's average returns can be interpreted as average return on individual anomalies that were historically significant. The average returns are expected to be positive if there is any persistence in returns on the anomalies. The significant anomalies are selected once a year, at the end of June. Only anomalies that are published by the time of selection are considered. Green, Hand, and Zhang (2017) documented a significant drop in performance of all anomalies in the US after 2003. A similar drop is observed on the portfolio-mixing strategy and it's average annualized return drops to less than 2% after accounting for transaction costs.

The strategy that synthesizes information from all the anomalies into one mispricing signal is studied next. The strategy first predicts next-month returns on individual stocks from their past characteristics (cross-sectional quantiles of the anomalies). Investment portfolios are then constructed by buying stocks in top decile of the predicted returns and short-selling stocks in the bottom decile of the predicted returns. Historical relation between the past characteristics and future returns is estimated on the past data. The next month returns on individual stocks are predicted from the latest characteristics. The historical relationships are typically linearly approximated using Fama and MacBeth (1973) least squares regressions in the academic literature, as in Lewellen et al. (2015). Gu et al. (2018) showed that machine learning methods can significantly outperform the linear approximation in the US. The use of machine learning methods is extended here from the US to international markets. The least squares regressions are compared to gradient boosting regression trees, random forest, and neural networks. The machine learning methods lead

[1] See Asparouhova, Bessembinder, and Kalcheva (2010) for description of the effect of microstructure noise.
[2] It is later shown that the results do not depend on the 5% significance level.

to significant gains in performance of the mispricing strategy in all the regions.

Value of international evidence for the prediction of out-of-sample returns on the anomalies is evaluated. Hou et al. (2017) and Harvey, Liu, and Zhu (2016) showed that many anomalies cannot be replicated and many others are significant only due to the in-sample data snooping. New anomalies are discovered using the same historical datasets in the US, which can lead to false positive discoveries. International data provides new information with respect to the US and it could therefore limit the number of false discoveries.[3] International data also increases sample size which in turn leads to more powerful statistical tests. One problem could be that some anomalies are specific to the US as they depend on the local institutional setting. For example, accruals depend on country-specific accounting rules. The institutional uniqueness then limits the value of data outside the US for predictions in the US.

Past performance of the individual anomalies from the US is the strongest predictor of their future performance in all the regions. There is also some evidence that past performance of the individual anomalies outside the US helps to predict their future performance in the respective regions but not elsewhere after accounting for the past performance in the US. Similar conclusions are also valid for the mispricing strategy. There is only a little gain in performance of the mispricing strategy in the US when the strategy's estimation sample is extended from the US stocks to international stocks. The profitability of the mispricing strategy in the other regions, however, improves when the estimation sample is extended from the US stocks to stocks in the respective regions. Mispricing of stocks estimated on historical data in the US captures most of predictability of stock returns outside the US.[4]

Marginal value of new anomalies for out-of-sample predictions after accounting for the already published anomalies is evaluated. Most of the widely accepted risk factors have been published before 1995. Examples include size and book-to-market ratio in Fama and French (1992) and momentum of Jegadeesh and Titman (1993). The new discoveries should therefore have lower marginal explanatory power over time as the strongest predictors of stock returns have been already revealed. It is also possible that the vetting procedure that authors have to undergo during the publishing process limits these decreasing returns to the new discoveries. The value of recent anomalies is examined by comparing out-of-sample returns of the mispricing strategy that synthesizes anomalies published either before 1995, 2000, or 2005. There is a gradual increase in mean

[3] Note that many anomalies have been individually studied in the international markets. For examples of studies investigating cross-sectional predictability of individual signals outside the US see Chui, Titman, and Wei (2010), Barber, De George, Lehavy, and Trueman (2013), McLean, Pontiff, and Watanabe (2009), Rouwenhorst (1998), Lam and Wei (2011), Titman, Wei, and Xie (2013), and Watanabe, Xu, Yao, and Yu (2013). The goal here is not the study of performance of the anomalies outside of the US but rather the use of international historical performance of the anomalies to better select anomalies that are likely to outperform in the future.

[4] The role of international evidence for the mispricing signal is broadly related to variety of factor structures outside the US. The international evidence is likely to add little value if there is no proximity of factor structures across the regions. For examples of papers investigating factor structure of international returns see Fama and French (2012), Fama and French (2017), Rouwenhorst (1999), Griffin (2002), Griffin, Kelly, and Nardari (2010), Hou, Karolyi, and Kho (2011), and Bartram and Grinblatt (2018).

returns and Sharpe ratio on the mispricing strategy over 2005 to 2016 period, when the more recently published anomalies are added. Investors can therefore benefit from following the recent academic anomalies research.

Limits to arbitrage could explain the strategies' profitability and it might not be possible to invest into the mispriced stocks. Several robustness checks are therefore conducted. The returns on the long-short portfolios are decomposed into long-only and short-only components. It is often impossible to short-sell certain stocks due to insufficient supply of borrowable shares. Both the long-only and short-only legs of the mispricing strategy, however, offer an investment opportunity with respect to returns on the market. Short-selling constraints cannot therefore fully explain the profitability. Transaction costs on the investment strategies are studied next. It is concluded that both the portfolio-mixing strategy and the mispricing strategy remain profitable after the transaction costs.[5]

The focus of this study is the closest to Jacobs and Müller (2017c) and Jacobs and Müller (2017a) who analyzed returns on anomalies outside the US. This study is, however, different in many aspects. Firstly, it focuses on liquid universe of stocks which should make the results more relevant to any investor. Secondly, the role of international evidence in the strategies is investigated. Jacobs and Müller (2017c) and Jacobs and Müller (2017a) focused solely on strategies that were using data in the respective regions without evaluating the possible benefits of using the global data to predict future returns. Thirdly, the prediction methods differ. The introduction of advanced machine learning techniques significantly improves the out-of-sample fit of the predictions in this study.

The study is the closest in methodology and application of machine learning techniques to Gu et al. (2018) who, however, focused solely on the US. Gu et al. (2018) in other respect, differ from this study with their focus on full universe of stocks, which has profound effects on their conclusions. The most important anomalies in their estimation are liquidity, size, and return over the past month (short-term return reversal). Asparouhova et al. (2010) argue that these variables are connected to future returns mainly through microstructure biases and have nothing to do with true predictability of stock returns that is of interest to investors.[6] The machine learning methods were built to find all patterns in the dependent variable and this leads to sub-optimal outcome when predicting stock returns on illiquid stocks. Focus on large cap universe helps to address these concerns. Secondly, a large difference with respect to Gu et al. (2018) is that this study allows

[5] Novy-Marx and Velikov (2015) studied transaction costs on a range of anomalies in the US and concluded that the transaction costs are important mainly for high-turnover anomalies whose returns net of transaction costs often turn negative. Frazzini, Israel, and Moskowitz (2012) demonstrated that real-life transaction costs for large portfolio managers are much lower than assumed by academics. In particular, returns on momentum and value style premia survive transaction costs and have large investment capacity. The transaction costs can be further lowered by appropriate optimized portfolio rebalancing.

[6] See Roll (1984) for a simple model decomposing stock returns into microstructure noise and changes in true prices.

only already published anomalies to enter predictions in each year. That is, the information set of existing anomalies was available to investors by the time they would make a decision of where to invest their money. Ignoring this assumption can lead to illusory profits that cannot be obtained in practice.

The contributions of this study are multiple. Firstly, the role of international evidence for predictions of future returns on individual stocks is evaluated. Most of academic anomalies research focuses solely on the US and benefits of international evidence have not been systematically studied before. It is shown that training sample outside the US does not largely improve forecasts of expected returns on the individual stocks in the US. Secondly, the marginal value of recent anomalies, while controlling for the well established anomalies, is evaluated. It is shown that the recently published anomalies are providing new information about the cross-section of stock returns.

I. Data and Methodology

A. Data

The source of accounting and market data for the US is Merged CRSP/Compustat database from Wharton Research Data Service (WRDS). The sample spans 1926 to 2016 period and contains all New York Stock Exchange (NYSE), Amex, and NASDAQ common stocks (CRSP share code 10 or 11). The returns are adjusted for delisting following guidance in Hou et al. (2017).[7]

The international data is sourced from Reuters Datastream. It is filtered following Ince and Porter (2006), Lee (2011), and Griffin et al. (2010). The procedure comprises of manually checking names of the shares in the database for over 100 expressions describing their share class. Only the primary quotes of ordinary shares of the companies are retained, with few exceptions where fundamental data in Datastream is linked to other share classes.[8] Real Estate Investment Trusts (REITs) are excluded from the sample. All the international returns and financial statements in this study are converted to US dollars. The daily returns are deleted for days when the stock market was closed in a given country. The quality of data is further improved with procedures described in Tobek and Hronec (2018) and covered in Appendix A. Tobek and Hronec (2018) study implications of the choice of fundamental database on the measurement of performance of individual fundamental anomalies. They show that statistical significance of the individual anomalies varies across Datastream and Compustat. The research inference can therefore change when a different

[7]If the delisting is on the last day of the month, returns over the month are used. The relevant delisting return is then added as a return over the next month. Delisting return (DLRET) from monthly file is used if it is not missing. $(1 + ret_{cum}) * (1 + DLRET_d) - 1$ is used if it is missing, where ret_{cum} is cumulative return in the given month of delisting and $DLRET_d$ is delisting return from the daily file. Lastly, the gaps are filled with $(1 + ret_{cum}) * (1 + DLRET_{avg}) - 1$, where $DLRET_{avg}$ is average delisting return for stocks with the same first digit of delisting code (DLSTCD).

[8]The description in Griffin et al. (2010) on classification of common shares is followed.

fundamental database is used. The differences across the databases are mainly due to imperfect historical fundamental coverage. Studies of aggregated performance of anomalies, however, do not suffer from these problems. Analysis in this study is therefore not impacted.

The sample includes 23 developed countries. The countries are sorted into 4 regions: the USA; Europe (E) - Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom; Japan (J); and Asia Pacific (AP) - Australia, New Zealand, Hong Kong, and Singapore.

Another important source of data for the anomalies is Institutional Brokers' Estimate System (I/B/E/S) which is obtained from WRDS. I/B/E/S is merged on Datastream directly as it is one of databases provided by Thompson Reuters and Datastream includes the respective tickers in its static file. The merger with CRSP is done indirectly through CUSIPs. The databases are merged on 8 digit CUSIP and then on 6 digit CUSIP if unsuccessful. The success of the merger is checked manually by comparing quoted tickers on the exchanges and names of the companies. All the variables in I/B/E/S are transformed to US dollars with original Reuters exchange rates which are provided by WRDS.

This study focuses mainly on the most liquid universe of stocks. The liquid universe category covers only stocks that are both (a) within the top 90% of the overall capitalization of all stocks in each region at the end of previous June and (b) within the top 90% of the overall dollar trading volume over the previous 12 months of all stocks in each region. The restriction on capitalization in the US roughly corresponds to 50% percentile of the largest stocks on NYSE. We also use milder all-but-micro-caps restriction in some parts of the analysis where the stocks are required to have capitalization larger than the bottom decile at NYSE at the end of previous June. This further capitalization constraint is also enforced for the liquid universe category to guarantee that the stocks outside the US are not only liquid with respect to other stocks in the region but also with respect to the stocks in the US. All the stocks are further required to have price larger than $1 ($.1 for Asia Pacific) at the end of the previous June.

Table I shows average, minimum, and maximum number of stocks in the cross-section of the individual regions. Full sample category includes all the available stocks without any restrictions. There are on average only about 500 stocks in the US that satisfy the criteria for the liquid universe. The average number of stocks satisfying the criteria in even smaller in the other regions. Average capitalization of stocks in the liquid universe after July 1995 is $24 billion in the US, $21 billion in Europe, $9 billion in Japan, and $11 billion in Asia Pacific. Average size of the stocks in the sample is therefore balanced over the regions.

Table I
Number of Stocks in the Cross-section

	Full sample			All-but-micro-caps			Liquid Universe		
	mean	min	max	mean	min	max	mean	min	max
Asia Pacific	2430	1012	3706	551	321	896	132	71	238
Europe	5194	4440	6121	1976	1410	2945	350	208	826
Japan	3141	2074	3678	1541	1030	2313	331	208	744
USA	4768	1993	7525	2340	1234	3852	495	263	829

B. Anomalies

The sample includes 153 anomalies published in academic studies. The full list of the anomalies is provided in Appendix B and their detailed description in the online appendix. Anomalies that have been described in McLean and Pontiff (2016), Hou et al. (2017), or Harvey et al. (2016) are primarily selected. The study focuses only on anomalies that are valid in the cross-section of stocks so that long-short portfolios can be formed out of them. Any anomalies that are specific to the US, and which cannot therefore be constructed outside the US, are excluded.[9] Fundamental signals are updated annually at the end of every June using financial statements from financial years ending in the previous calendar year.[10]

Some anomalies, such as Herfindahl Index of Hou and Robinson (2006), require classification of industries for individual firms. The choice in the original papers is mostly with respect to Standard Industrial Classification (SIC). Third level Datastream classification, sorting industries into 19 groups, is applied here instead. The larger industry groups should make the results more robust and consistent across the data vendors. The industry classification in Datastream is available only from the static file, which means that only the latest values are available. Data vendors may slightly differ in the classification of individual firms over time because the differences between individual SIC categories are often subtle. A translation table between SIC classification and the Datastream classification is provided in the online appendix.

There are 93 fundamental, 11 I/B/E/S, and 49 market friction anomalies in the sample. The anomalies come almost exclusively from the top finance and accounting journals. Figure 1 graphs number of the published anomalies over time. The second line is capturing number of anomalies whose in-sample period in their respective studies has ended. The number of anomalies has been gradually increasing over time without any apparent jumps.

[9]This includes anomalies: based on quarterly fundamental data since there is only short coverage internationally; connected to hand collected data in the US such as IPOs, SPOs, and mergers; requiring segment information and NBER data; and that are institutionally specific, such as, share turnover or effective tax rate. Some fundamental anomalies could not be implemented in Datastream as the required items are missing there.

[10]Section A documents that the annual refreshing of fundamental signals provides very similar results to monthly refreshing.

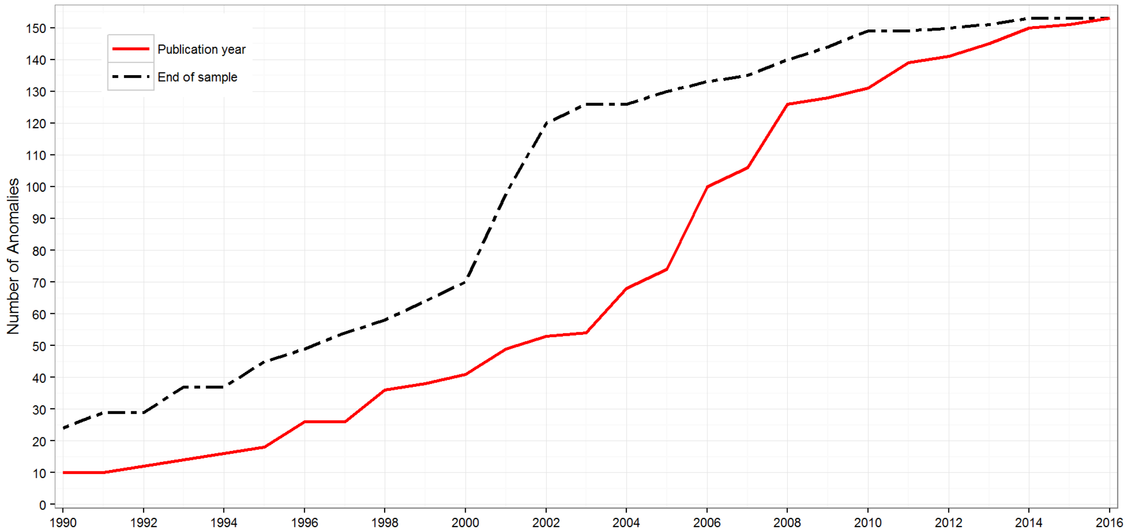


Figure 1. Number of the Published Anomalies Over Time.

C. Portfolio-mixing Strategy

This section describes the portfolio-mixing strategy that equally weights returns from the portfolios on individual anomalies. It serves as a benchmark for the more complicated mispricing strategy described in the next section. The strategy is especially useful when studying the role of international evidence in the selection of quantitative strategies that outperform out-of-sample as it can be understood as a combination rule of multiple quantitative strategies based on the past evidence. Portfolio construction for the individual anomalies is first described and the logic for how the individual portfolios are combined is discussed next.

C.1. Portfolio Construction for the Individual Anomalies

The portfolios are constructed on the liquid universe of stocks. The focus on liquid universe should make the findings more realistic to someone trying to trade the anomalies. The stocks with small capitalization (micro-cap) account for only a small fraction of the overall capitalization of the market and often cannot be traded at significant volumes due to their high illiquidity. Tobek and Hronec (2018) document that the fundamental coverage of micro-cap stocks outside the US is very problematic in Datastream and the imperfect coverage can introduce huge biases into the analysis. Both equal-weighted and value-weighted returns are always provided. The preference should be given to interpretation of the value-weighted returns since they do not suffer from the market microstructure biases documented in Asparouhova et al. (2010). These biases can be substantial and can heavily influence the analysis.

The portfolios on individual anomalies start in July 1963 in the US and July 1990 in Europe, Japan, and Asia Pacific.[11] The period before 1963 in the US is omitted due to the quality of returns and number of available stocks in CRSP is very low during that time. The fundamental coverage of stocks in Compustat is also very low which makes the construction of majority of the anomalies impossible. Further restrictions of the sample of stocks, based on industries, age of the firms, and the length of history of the firms' fundamental data, follow the original studies when constructing portfolios on the individual anomalies. The original studies are also followed regarding rebalancing period of the portfolios so that most of anomalies in I/B/E/S and market friction categories are rebalanced monthly, whereas, fundamental anomalies are mostly rebalanced annually at the beginning of every July. The zero-cost long-short portfolios on the individual anomalies are constructed by buying stocks in the top decile of the signals and shorting stocks in the bottom decile of the signals.[12]

Table II
Average Time-series Correlations of Returns on Portfolios Created for the Individual Anomalies Across the Regions.

	USA	E	J	AP
USA	1.000	0.239	0.105	0.120
E	0.239	1.000	0.126	0.122
J	0.105	0.126	1.000	0.094
AP	0.120	0.122	0.094	1.000

Table II presents average of time-series correlations of returns on the long-short portfolios created from identical anomalies across the different regions. The anomalies are not closely correlated across the regions. The international evidence should therefore be very useful as it can serve as an independent source of information for stock return predictability.

C.2. Combining Individual Portfolios into One Strategy

The portfolios on individual anomalies are combined into one meta-portfolio through a simple strategy. The portfolio-mixing strategy selects all the anomalies whose portfolio returns are significant at 5% level and equally weights them into a single portfolio. The selection is repeated at the end of every June from 1995 to 2016. Many of the published anomalies cannot be replicated on the liquid universe of stocks and the selection based on historical significance guarantees that only robust strategies are used. Significance of the anomalies is determined based on returns available up to the given June. Only anomalies published by the given June are considered. The significance is determined based on p-values that are adjusted for heteroskedasticity and auto-correlation for

[11] International studies using fundamental data, such as Fama and French (2017), usually start in 1990. The reason for this is that there is an insufficient fundamental coverage before that.

[12] The zero-cost portfolios are preferred since some annually rebalanced anomalies experience lower than -100% return during some years which creates problems with the definition of return in terms of relative change in value of the invested money with respect to the previous month. It would be necessary to introduce leverage constraints which would unnecessarily complicate the analysis.

up to 12 lags.

The equal-weighting of portfolios on individual anomalies adds robustness to the strategy. It could be beneficial to use information of historical covariance structure between the strategies. DeMiguel, Garlappi, and Uppal (2007), however, show that 1/N weighting provides a very robust performance out-of-sample and no other simple weighting strategy is able to beat it.

D. *Mispricing Strategy*

The focus has so far been on portfolio level analysis of the individual anomalies. The rest of this section covers the strategy that shrink all the anomalies into a single mispricing signal ("mispricing strategy"). Lewellen et al. (2015) defined the prediction problem as follows: the goal is to devise a forecasting method that predicts which stocks are likely to have the highest returns in the next month and which the lowest based on stock characteristics (the cross-sectional anomalies). To do this, monthly returns on individual stocks are regressed on their past characteristics. The future returns are then predicted from the latest available characteristics. The regressions are estimated by pooling all the available stock returns up to the date of portfolio formation. The past characteristics have to be available before the start of measurement period of the returns. The characteristics are normalized to their cross-sectional quantiles within each region to reduce problems with outliers.

To summarize, the following equation is estimated

$$r_{it} = f(x_{i,t-1,1}, x_{i,t-1,2}, ..., x_{i,t-1,M}) + \epsilon_{it} \tag{1}$$

where r_{it} is return on stock i in month t and $x_{i,t-1,1}$ is cross-sectional quantile of a given anomaly (characteristic) for the stock i available before the start of month t. The returns are demeaned by subtracting average cross-sectional returns in every region-month. A simpler case with linear $f()$ is first covered. It is then extended to a more general structure using machine learning. The machine learning exercise follows Gu et al. (2018) who applied a suite of standard machine learning algorithms and showed that they outperform the linear models in the US. Readers are referred to Gu et al. (2018) or any advanced machine learning textbook for a detailed theoretical description of the machine learning methods and only basic definitions are covered here.[13]

The more complicated machine learning methods require a large training sample to work properly. The liquid universe of stocks as defined in Section I.A can be too small for the estimation purposes. All the mispricing strategies are therefore estimated on the more numerous all-but-micro-caps sample of stocks, which is also defined in Section I.A.[14]

[13] See, for example, Friedman, Hastie, and Tibshirani (2001) for the textbook treatment.

[14] Table IA2 in the Online Appendix documents that the impact of this choice is only tiny for all the estimation

The machine learning methods have both some benefits and some negatives. They provide better out-of-sample forecasts through limitation of in-sample over-fitting. They also allow for a very general interaction between the explanatory variables. This general form, however, makes the fitted models hard to estimate and the estimates hard to interpret due to the black-box approach. The intractability of the estimates is not a large concern in this study since even the linear method becomes intractable given the number of exogenous variables. The main metric of this study is out-of-sample performance and not the interpretation of the estimated parameters, which is in line with the optimization objective of the machine learning methods.

The machine learning methods usually depend on some pre-specified meta-parameters. This study follows the common approach in machine learning literature to choose the meta-parameters in data-dependent way through three-fold cross-validation (CV). The CV splits the historical sample into pairs of mutually exclusive validation samples and training samples. The model is estimated on the training sample with various meta-parameters and its performance is captured on the validation samples. The meta-parameters, maximizing the performance over all the validation samples, are then selected for the estimation. The CV splits divide the historical sample into three consecutive parts with similar length.[15]

D.1. Weighted Least Squares

The benchmark model uses weighted least square estimation for linear approximation of the relationship in equation (1). That is, a weighted least square regressions of the stock returns on the rescaled characteristics is estimated,

$$r_{it} = \beta_0 + \beta_1 x_{i,t-1,1} + \beta_2 x_{i,t-1,2} + ... + \beta_M x_{i,t-1,M} + \epsilon_{it} \tag{2}$$

where the weight on individual observation is the inverse of number of stocks in the each time period and region. The weights are introduced to give equal importance to the each time period. The weighting makes the moment conditions equivalent to Fama and MacBeth (1973) regressions in Lewellen et al. (2015). The linear specification has already been applied in international context in Jacobs and Müller (2017c) and Jacobs and Müller (2017a). It is therefore selected as a benchmark for the more complicated machine learning methods.[16]

methods apart for neural networks which benefit from the larger sample.

[15] The sample splits for the initial historical sample period 1963 - 1995 are, for example, 1973 and 1984. The pairs of training and validation samples are then [1963 - 1984, 1985 - 1995], [1963 - 1973 plus 1985 - 1995, 1974 - 1984], and [1974 - 1995, 1963 - 1973].

[16] Capitalization-weighted regressions as in Green et al. (2017) have been also tried. The capitalization-weighting puts lower weight on small cap stocks and is more suited for value-weighted portfolios. The weighting did not outperform the selected method and the results are therefore not reported here.

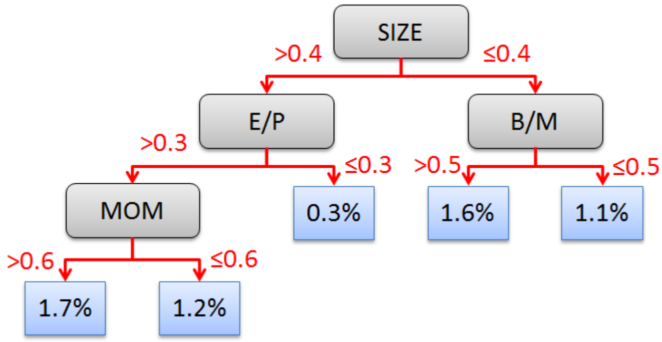


Figure 2. Decision Tree.

D.2. Penalized Weighted Least Squares

The linear regression model with many explanatory variables can overfit the realization of past data since it has many degrees of freedom. One way how to reduce the overfitting is to introduce L1 and L2 penalties on the coefficients during the estimation. The penalties are chosen by the three-fold cross-validation. The cross validation mostly selects only L1 penalty. The case with just L1 penalty is denoted least absolute shrinkage and selection operator (LASSO) and was introduced in Tibshirani (1996).

D.3. Random Forest

The regression tree family of methods is easy to estimate and requires a few specified meta-parameters. One such tree is depicted in Figure 2. The decision tree consists of nodes (the round-edged boxes) and outcomes (sharp-edged boxes). The outcomes are in percent return per month.[17] The tree starts with a decision whether a given stock is within the smallest 40% of stocks in the cross-section. The decision can then continue to the split based on the book to market ratio. The depicted tree is of depth 3, which is the maximum number of nodes in the longest branch. The tree allows for arbitrary cross-effects between the variables up to the (depth - 1) degree. This study deals mainly with relatively shallow trees. The shallow trees are nonetheless able to capture various important interactions between the explanatory variables. Random Forest and Gradient Boosting Regression Trees are based on a combination of the individual trees. These methods cannot be easily visualized but they lead to a better out-of-sample forecasting performance relative to simpler regression trees.

Random forest is one of the most widely used ensemble tree method. It combines forecasts from the individual decision trees that are based on subsamples of the training data. Explanatory variables are also subsampled in the individual trees to increase variety among the individual forecasts. Random forest is frequently among the top 10% of best performing machine learning

[17]The numbers are arbitrary and do not reflect real data.

methods in various competitions and it is therefore a very robust method that is powerful in most of the settings. It requires only few specified meta-parameters. The specification of the meta-parameters is furthermore not very important for its performance. It can therefore be used almost out-of-box. This is a large benefit with respect to neural networks where performance heavily depends on specification of the model. The largest downside is that its estimates is time consuming.

The results in this study are based on a combination of 500 trees. The trees use randomly selected 50% of the overall training observations and square root of the overall available explanatory variables. Minimum node size is chosen to be 0.1% of all the training observation to leave the method completely meta-parameter free. The 0.1% is large enough to limit over-fitting but small enough to allow the method to approximate the true expected returns on stocks.[18]

D.4. Gradient Boosting Regression Trees

Gradient boosting regression trees (GBRT) of Friedman (2001) rely on a different way of combining the regression trees than random forest. All the trees in random forest are chosen independently, whereas, they are selected in a dependent fashion in GBRT. The idea is to estimate a tree and use only a fraction of its fit for forecasts. The next iterations then proceed on residuals of the dependent variable after removing the fraction of the fitted values in the previous iteration. Shrinkage of the individual predictions guarantees that the learning can correct itself if the fitted values are selected suboptimally in some iterations. The fraction of individual predictions that is retained for the forecast is called a learning rate. Number of the learning iterations, given the learning rate, then determines how closely the particular realization of the sample from the whole population (the training sample) is over-fitted. A selection of fewer iterations reduces the risk of over-fitting (estimation error) but decreases the overall fit of the estimation (i.e. introduces an approximation error). It is therefore important to select the number of iterations with optimal estimation and approximation error trade-off. One way to do this is to rely on a cross-validation. The method requires a specification of learning rate, number of iterations (trees), and maximum depth of the trees.

The analysis in this study is conducted with a fast version of the gradient boosting - extreme gradient boosting (XGBOOST) of Chen and He (2017). The reason for this is that it is ten times faster to estimate and thus requires far less computational power. Gu et al. (2018) benchmarked the different machine learning methods and only neural networks provided significantly better forecasts than GBRT. GBRT is therefore a good candidate for the empirical application and it captures most of the gains from the machine learning methods over the standard finance methods.

[18]Ignoring this this parameter completely, and leaving unlimited note size, leads to almost identical results. It is thus not an important assumption.

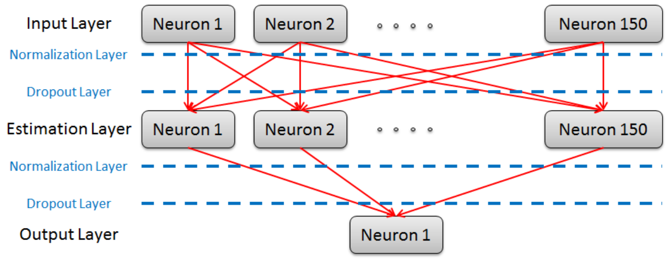


Figure 3. Neural Network.

That is why GBRT is used to examine the benefits of international training sample in section III and the benefits of recent anomalies in section IV.

The specification of the GBRT is set as follows: the maximum depth of the trees is determined by a cross-validation. Depth of up to 9 nodes is considered. Gu et al. (2018) showed that cross-validation selects similar values in their analysis. The learning rate is set to 10%.[19] Number of iterations is again determined via the three-fold cross-validation.

D.5. Neural Networks

Arguably the most powerful machine learning method of today is (deep) neural networks. Gu et al. (2018) show that they outperform any other method if they are optimally specified. The neural networks are a very flexible tool that encompasses many specifications.[20] The flexibility is also their largest disadvantage as it requires a long experimentation and possible over-fitting of the sample.[21]

Sequential neural networks consist of layers of neurons with information flowing between the layers in only one direction, from input layer to output layer. The information is fed in batches consisting of n sample points. Processing of the full training sample is called an epoch. The speed of change in estimated parameters with new processed batches is determined through the learning rate. It is often an advantage to slow the learning rate over time to allow for finer details to be captured. The neural networks are estimated with back-propagation and stochastic gradient descent.

Figure 3 plots specification of the neural network in this study. It is based on three layers. The initial layer has 150 neurons. The second hidden layer also has 150 neurons. The last output layer only has one neuron. The first two layers use a rectified linear unit (ReLU) activation function while the last layer uses a linear activation. Input into each layer is batch normalized. The network

[19] Experimenting with the learning rate did not lead to any increase in the predictive power. There is an extreme amount of noise in the financial data and slower learning is thus not necessary.

[20] A linear regression is the simplest specification.

[21] The over-fitting should be a large cause of worry and all results based on neural networks should be taken with a grain of salt. The tree-based methods work well out of box even with default setting but neural networks require a long fine tuning. The fine tuning will translate into problematic performance out-of-sample of this study.

is regularized with dropout layers where output of fifteen randomly selected neurons is dropped in the first and the second layer in each epoch. Early stopping callbacks then provide further regularization and stop the learning process once the mean squared loss stops improving in the validation sample in four consecutive epochs. Another callback reduces the learning rate when the mean squared loss stops improving from one epoch to another.

The final forecast is produced from a combination of three estimated neural networks with different initial random seeds. Each run also uses different validation-training sample splits to further increase variety over the forecasts. The combination forecast leads to a great improvement in the performance of the mispricing strategy based on the neural networks.

D.6. Portfolio Construction

The mispricing portfolios start in July 1995, unless stated otherwise. They are again long-short self-financing and are rebalanced every month. The long leg of the strategy buys stocks in the upper decile of the predicted next month's returns. The short leg of the strategy short-sells stocks in the bottom decile of the predicted next month's returns.

The portfolios are constructed based on sorts of the predicted returns in the individual regions. Global strategy invests into stocks from all the four regions. The global strategy is again based on stocks in the extreme deciles of the predicted returns in the individual regions.

The portfolio returns now also correspond to an investable strategy that holds $1 in cash, invests $1 in the stocks that are likely to have the largest return in the next month, and shorts $1 worth of stocks that are likely to have the smallest return in the next month. The portfolios are rebalanced to have an equal position in cash, long, and short leg of investment in the stocks at the beginning of each month.

E. Liquidity Measures

Liquidity costs on the strategies are studies with several liquidity proxies. The proxies are: VoV(% Spread) of Fong, Holden, and Tobek (2017), Gibbs proxy of Hasbrouck (2009), and closing quoted spread proxy of Chung and Zhang (2014). They are defined in detail in Appendix C.

The proxies were selected to capture a fixed component of transaction costs and ignore variable component that measures price impact of larger orders. The variable component is very volatile and depends on the precise trade execution algorithm of each asset manager. The large capitalization universe of stocks reduces concerns about the variable component and it should be possible to avoid any execution costs altogether through the use of limit orders.

All of the proxies have some missing observations. The missing observations are backfilled from the other proxies. Quoted spread is used first for the backfilling, followed by VoV(% Spread), and the remaining missing observations are backfilled with Gibbs proxy. Less than 0.02% of the observations is missing in all the three proxies and these observations are filled by 5% costs.

II. Profitability

A. *Portfolio-mixing Strategy*

The portfolio level analysis of the individual anomalies is a good starting point as it provides a simple indication of out-of-sample profitability of the anomalies. The more complicated method, that synthesizes information embedded in the individual anomalies to one mispricing signal, is just a refinement of this simple strategy.

Table III presents returns on the portfolio-mixing strategy that invests equally in all the portfolios on anomalies that have significantly positive returns at 5% significance level as described in Section I.C.2. That is, it corresponds to a setting where someone is following anomalies research, replicates the published findings, and equally invests into all published anomalies that he was able to replicate on the liquid universe of stocks. The performance of the portfolio-mixing strategy is followed in all the regions. The out-of-sample forecasts begin in July 1995. Global strategy equally invests in the portfolio-mixing strategy in the four developed regions.

Table III
Out-of-sample Performance of the Portfolio-mixing Strategy

The table shows returns of the strategy that equally invests in all the anomalies that are significant in the US at 5% significance level as described in Section I.C.2. The significant anomalies are selected once a year, at the end of June. Only anomalies that are published by the time of selection are considered. The reported returns are for July 1995 to December 2016 period and are in percentage points.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Mean	0.174	0.297	0.001	0.663	0.284	0.301	0.180	0.253	0.882	0.404
Sharpe Ratio	0.227	0.484	0.002	0.695	0.566	0.387	0.270	0.198	0.816	0.598
Skewness	0.083	-0.085	-1.885	-1.087	-0.436	0.356	0.197	-0.046	1.871	1.358
Kurtosis	9.963	9.230	14.68	13.16	6.920	6.481	9.243	25.09	16.59	22.41
Max Drawdown	-29.40	-17.96	-27.63	-27.43	-12.95	-18.12	-26.33	-61.07	-17.35	-20.79

The portfolio-mixing strategy is not statistically significant in the US for both equal-weighted and value-weighted returns over 1995-2016 period and Sharpe ratio is also low there. The profitability is sometimes higher in the other regions. The strategy is the most profitable in Asia Pacific. The returns are higher outside the US despite the fact that the anomalies have been chosen in the US without any regard for evidence from the other countries. The anomalies documented in academic literature in the US are therefore successful in capturing risk premia outside the US. Diversification among the regions also provides some benefits. The global strategy has Sharpe

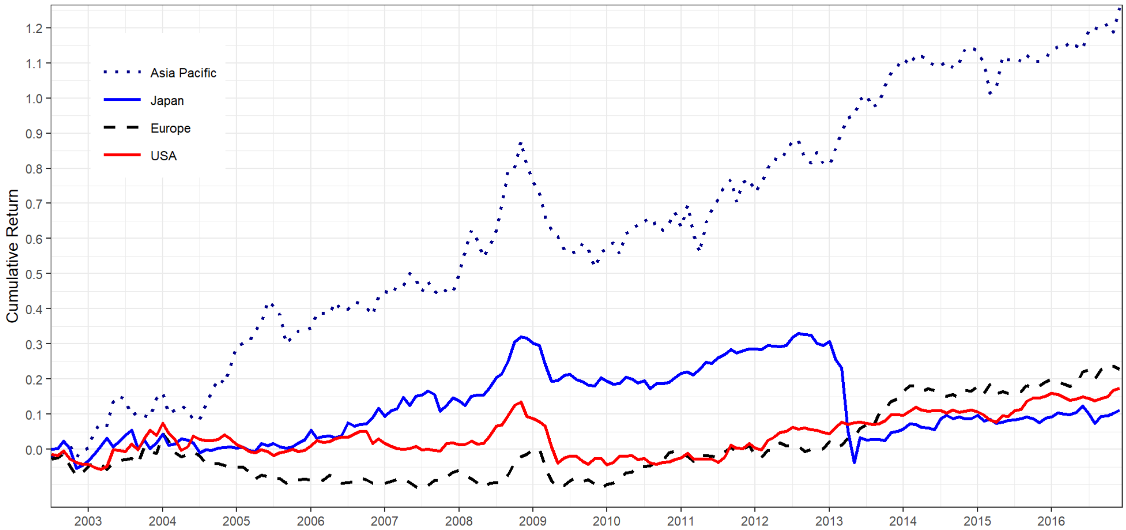


Figure 4. Cumulative returns on the Portfolio-mixing Strategy.

ratio close to 0.6.

Maximum drawdown (DD) is defined as

$$\min_{s>t} 100 * (P_s/P_t - 1) \tag{3}$$

where P_t is market value of all assets held in the strategy at time t. That is, DD is the largest relative drop in value of the invested money over the 1995 to 2016 period. DD is the smallest in Asia Pacific regions for value-weighted returns, which is in line with the highest returns and Sharpe ratio there. It is, nonetheless, also small in other regions, except for Japan.

Green et al. (2017) showed that the profitability of all anomalies has decreased significantly after 2003. The same decline in profitability is documented in Figure 4. The figure presents evolution of cumulative returns on the portfolio mixing strategy since June 2002. The profitability of the individual anomalies in the US has dropped to the point that they yielded only about 20% in this whole period. The strategy was more profitable in other regions.

The portfolio-mixing strategy relies on a specific threshold for the decision whether to include a given anomaly in the mix. Figure 5 documents that the results are robust to the choice of this threshold. The figure shows annualized mean returns and Sharpe ratios for the portfolio-mixing strategy that equally invests into all anomalies whose historical returns have t-statistic larger than threshold specified at x-axis. Mean returns are increasing with the threshold in all the region. The mean return on the anomalies are therefore larger the more significant they were historically. Sharpe ratio of the portfolio-mixing strategy does not depend that strongly on the significance threshold.[22]

[22]Note that there are only few anomalies with t-statistic larger than 2.5 and the results become unstable after

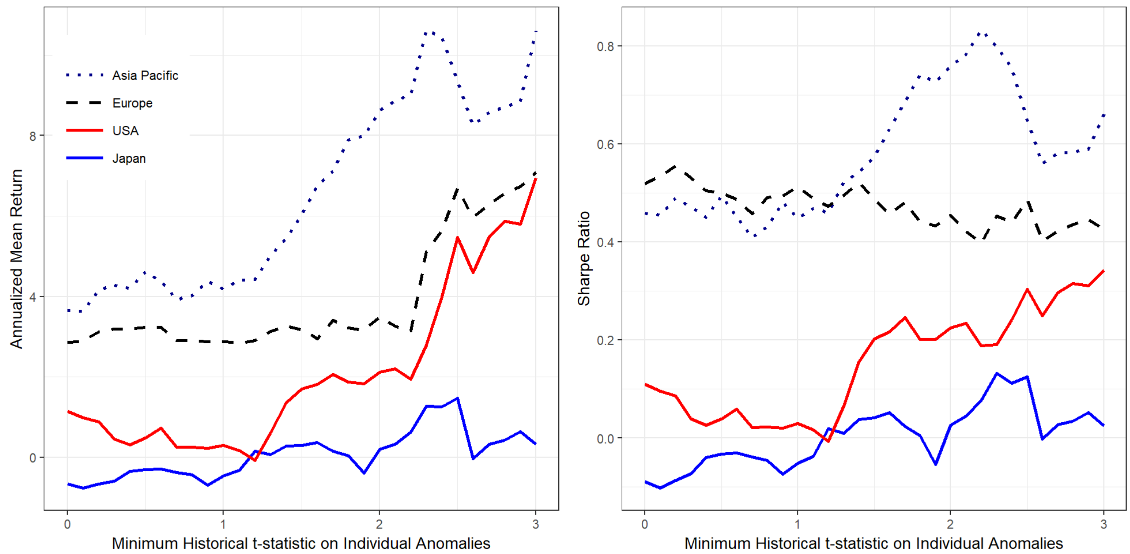


Figure 5. Annualized Mean Returns and Sharpe Ratios on the Portfolio-mixing Strategy Depending on Significance Threshold for Individual Anomalies.

To conclude, returns on the anomalies in all the regions are positive, which suggests that it is profitable to invest in the anomalies before transaction costs.

B. *Mispricing Strategy*

Performance of the mispricing strategy is examined next. Jacobs and Müller (2017a) showed that the mispricing strategy estimated with least squares leads to higher returns in both absolute term and on risk adjusted basis relative to mixing of portfolios on individual anomalies. Gu et al. (2018) then documented that the more sophisticated machine learning methods provide higher out-of-sample predictability relative to least squares. The machine learning methods are extended to the international sample to determine whether their benefits persist outside the US.

Table IV presents mean returns on portfolios created based on the mispricing strategy. The regressions of stock returns on their characteristics are fit on data available up to June every year and the future stock returns are then predicted with the latest available characteristics for each of the next 12 months. The regressions are estimated with least squares, penalized least squares, random forests, gradient boosting regression trees, and neural networks. The estimates in table IV are based on the US data from July 1963. The long-short decile portfolios, that invest into stocks in the top decile of the predicted future returns and short-sell stocks in the bottom decile of the predicted returns, are then created. The reported returns on portfolios are in percent per month and are from July 1995 to December 2016.[23]

that.

[23]Table IA2 in the Online Appendix provides comparison between estimation of the mispricing strategies on the all-but-micro-caps sample and the liquid universe of stocks. Equal-weighted portfolios estimated on the liquid universe slightly underperform those estimated on all-but-micro-caps. The underperformance is the largest for the

Table IV
Performance of the Mispricing Strategy Estimated in the US

The table shows out-of-sample performance of the mispricing strategy as defined in Section I.D. It is based on long-short decile portfolios from the strategy that combines all the available anomalies through predictive regressions of individual stock returns on transformed characteristics. The estimation methods are least squares, penalized least squares, random forests, gradient boosting regression trees, or neural networks. That is, pooled regressions of monthly stock returns on cross-sectional quantiles of their characteristics observable before each month start are estimated and future returns from the latest available characteristics are predicted. The value-weighted or equal-weighted long-short portfolios are constructed by buying stocks in the top decile of the predicted next month returns and shorting stocks in the bottom decile of the predicted next month returns. The regressions are rerun at the end of each June with only those anomalies that have been published by that time. The out-of-sample performance is observed in the US, Europe, Japan, and Asia Pacific. The training sample spans July 1963 to December 2016 in the US and July 1990 to December 2016 in other regions. The regressions are estimated only on the past US data and the future returns are predicted in all the regions. The reported returns are for July 1995 to December 2016 period and are in percentage points.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
					Weighted Least Squares					
Mean	0.801	0.680	0.922	0.782	0.810	0.575	0.647	0.648	0.633	0.639
Sharpe Ratio	0.479	0.541	0.701	0.497	0.763	0.348	0.472	0.410	0.318	0.550
Skewness	-0.340	0.239	-0.425	-0.356	0.025	-0.121	-0.017	-0.688	-0.238	-0.041
Kurtosis	8.521	5.673	4.167	3.713	8.488	7.214	6.483	5.544	4.824	7.880
Max Drawdown	-64.70	-37.10	-43.36	-47.61	-43.51	-69.75	-34.16	-44.52	-49.86	-50.85
					Penalized Weighted Least Squares					
Mean	0.756	0.728	0.886	0.854	0.800	0.644	0.794	0.596	0.754	0.683
Sharpe Ratio	0.443	0.557	0.665	0.526	0.724	0.381	0.554	0.372	0.365	0.568
Skewness	-0.487	0.067	-0.620	-0.365	-0.263	-0.316	0.123	-0.692	-0.409	0.023
Kurtosis	8.703	6.662	4.540	3.834	9.301	7.297	7.579	5.715	5.080	9.111
Max Drawdown	-65.36	-35.46	-42.32	-48.50	-45.10	-68.02	-37.51	-49.44	-57.38	-49.59
					Gradient Boosting Regression Trees					
Mean	1.165	0.870	1.173	1.650	1.163	1.391	0.591	1.011	1.415	1.033
Sharpe Ratio	0.720	0.644	0.766	1.005	1.146	0.831	0.412	0.525	0.800	0.870
Skewness	0.319	-1.160	0.682	-0.437	-0.449	0.561	-1.314	0.800	-0.112	-0.433
Kurtosis	6.653	10.17	8.274	5.575	6.812	9.287	12.03	8.611	4.718	7.797
Max Drawdown	-38.31	-48.25	-34.37	-36.65	-27.45	-43.93	-42.31	-41.79	-39.58	-35.62
					Random Forest					
Mean	1.050	1.037	1.107	0.943	1.074	0.977	0.339	1.028	1.183	0.798
Sharpe Ratio	0.703	0.782	0.781	0.520	1.080	0.691	0.222	0.591	0.612	0.726
Skewness	-0.328	-1.132	-0.283	-0.789	-0.939	-0.594	-1.149	0.675	-0.062	-0.974
Kurtosis	5.989	9.399	5.558	7.201	7.191	4.951	12.27	8.613	5.855	7.549
Max Drawdown	-30.69	-48.18	-40.16	-46.87	-27.76	-30.59	-54.54	-42.12	-39.57	-31.17
					Neural Networks					
Mean	1.416	1.097	1.295	1.752	1.346	1.420	0.826	1.100	1.177	1.093
Sharpe Ratio	0.905	0.880	1.130	1.086	1.582	0.905	0.649	0.693	0.697	1.042
Skewness	-0.083	-0.082	-0.149	0.244	-0.310	-0.167	-0.470	0.629	0.638	-0.255
Kurtosis	7.316	4.827	4.446	5.091	5.304	6.432	7.050	10.37	5.075	6.806
Max Drawdown	-44.60	-33.93	-24.70	-38.10	-18.90	-48.11	-31.93	-37.09	-54.45	-33.25

Both, the tree based methods and neural networks, outperform simple least squares. In particular, gradient boosting regression trees and neural networks outperform least squares in all the regions for both mean returns and risk adjusted Sharpe ratios. The machine learning methods

most complex method - neural networks. Value-weighted portfolios are, however, more profitable when estimated on the liquid universe with the exception of neural networks. The estimation on the liquid universe leads to more homogeneous return predictions over the liquid stocks as equal-weighting leads to almost identical performance as value-weighting.

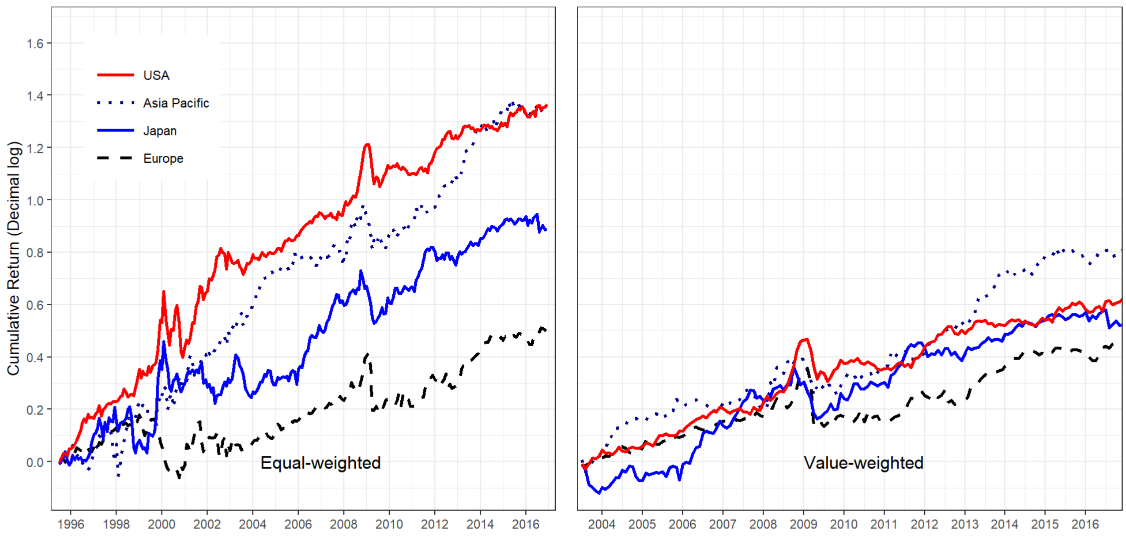


Figure 6. Cumulative Returns on the Gradient Boosting Regression Trees Mispricing Strategy. The figure shows cumulative returns for the mispricing strategy as described in Table IV that is estimated on the individual stocks from the US.

are therefore more powerful for stock return predictions outside the US as well as inside the US. The superior performance outside the US provides robustness to findings in Gu et al. (2018), who focused solely on the US. The average returns on the mispricing strategies are about 4 times higher than for the portfolio level strategy in the previous section.

Gradient boosting regression trees and neural networks also have the smallest maximum drawdowns and investing in them is therefore the least risky. Diversification over the four regions (in the global columns) further reduces the maximum drawdowns and increases the Sharpe ratios.

Figure 6 plots cumulative returns on the gradient boosting regression tree mispricing strategy in Table IV. The returns are presented in decimal logarithms and 1 on the left scale therefore corresponds to 1000% return on the initial investment. There is a small drop in profitability around 2003 in the US, which is in line with the evidence from portfolio-mixing strategy in Figure 4. The mispricing strategy is the least profitable in the European region.

B.1. Long-only and Short-only Components of the Strategy

Short-selling can be connected to large costs and sometimes even outright impossible. That is why it might not be possible to replicate the returns on the mispricing strategy in practice.[24] The long-short strategy in Table IV will now be decomposed into long-only and short-only components to determine the role of short-selling for the strategy's profitability. Table V decomposes the long-short returns separately for the individual machine learning methods. The long-only component

[24]Short-selling constrains should not be a large issue on our liquid universe of stocks. Andrikopoulos, Clunie, and Siganos (2013) showed that although some stocks cannot be short-sold in practice, focusing only on those that can be short-sold does not statistically diminish returns on 8 quantitative strategies in the UK. They also showed that short-selling costs are small at about 1% annually in the UK.

can be compared to equal-weighted and value-weighted returns on the whole market as defined by the liquid universe of stocks in Panel A.

Table V
Decomposition of the Returns on the Mispricing Strategy to Long-only and Short-only Components

The table shows returns of the mispricing strategy described in Table IV that is estimated on the individual stocks from the US. The returns on the long-short portfolios are decomposed to long-only and short-only components. Equal-weighted and value-weighted returns on the whole stock markets in the individual regions estimated on the liquid sample of stocks are also provided.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Panel A: Long-only Component of the Mispricing Strategy										
Whole Market										
Mean	0.829	0.739	0.300	0.737	0.655	0.786	0.668	0.222	0.806	0.617
Sharpe Ratio	0.507	0.445	0.184	0.376	0.439	0.609	0.442	0.145	0.471	0.477
Skewness	-0.656	-0.540	0.075	-0.565	-0.684	-0.651	-0.539	0.052	-0.492	-0.738
Kurtosis	4.391	5.215	3.260	5.405	5.447	3.883	4.249	3.103	4.707	4.384
Max Drawdown	-60.81	-63.61	-58.18	-64.72	-56.68	-51.41	-59.46	-65.98	-59.07	-54.54
Weighted Least Squares										
Mean	1.099	1.066	0.700	0.786	0.946	0.889	0.990	0.689	0.923	0.812
Sharpe Ratio	0.626	0.585	0.380	0.323	0.605	0.570	0.534	0.382	0.420	0.549
Skewness	-0.703	-0.611	0.072	-1.103	-0.811	-0.654	-0.535	-0.012	-0.814	-0.513
Kurtosis	5.328	5.045	3.908	8.174	5.229	5.225	5.024	3.970	8.804	4.554
Max Drawdown	-56.12	-62.66	-60.40	-76.62	-59.38	-48.56	-59.94	-63.29	-65.16	-52.87
Information Ratio	0.339	0.505	0.525	0.052	0.342	0.120	0.390	0.465	0.093	0.289
Penalized Weighted Least Squares										
Mean	1.036	1.063	0.643	0.848	0.913	0.867	1.064	0.614	1.057	0.794
Sharpe Ratio	0.595	0.594	0.352	0.343	0.592	0.563	0.582	0.347	0.480	0.552
Skewness	-0.837	-0.608	0.060	-1.094	-0.843	-0.749	-0.532	-0.063	-0.700	-0.606
Kurtosis	5.750	5.178	3.707	8.003	5.271	5.159	5.153	3.555	9.061	4.542
Max Drawdown	-56.87	-63.39	-59.58	-75.79	-59.31	-49.53	-58.57	-59.86	-67.70	-53.04
Information Ratio	0.251	0.480	0.466	0.108	0.300	0.093	0.471	0.405	0.201	0.265
Gradient Boosting Regression Trees										
Mean	1.235	1.154	0.676	1.414	1.078	1.367	0.986	0.653	1.396	1.084
Sharpe Ratio	0.569	0.654	0.357	0.600	0.629	0.684	0.586	0.360	0.625	0.650
Skewness	-0.338	-0.717	0.191	-0.530	-0.602	-0.020	-0.444	0.399	-0.485	-0.347
Kurtosis	6.051	5.357	3.818	5.681	4.314	6.596	4.746	5.293	5.455	4.035
Max Drawdown	-71.09	-63.32	-61.80	-63.02	-57.61	-65.67	-61.14	-73.43	-61.49	-62.35
Information Ratio	0.456	0.718	0.472	0.725	0.468	0.500	0.487	0.448	0.498	0.624
Random Forest										
Mean	1.127	1.275	0.577	0.971	0.994	0.951	0.968	0.620	1.003	0.868
Sharpe Ratio	0.527	0.709	0.315	0.396	0.585	0.523	0.553	0.327	0.446	0.547
Skewness	-0.985	-0.688	0.143	-0.501	-0.788	-0.975	-0.603	0.380	-0.316	-0.576
Kurtosis	6.545	5.300	3.468	5.211	4.452	6.248	5.125	4.849	4.467	3.825
Max Drawdown	-76.51	-62.27	-64.61	-75.60	-62.19	-69.23	-61.92	-74.02	-70.31	-63.97
Information Ratio	0.356	0.941	0.388	0.243	0.382	0.185	0.394	0.438	0.153	0.402
Neural Networks										
Mean	1.295	1.262	0.756	1.381	1.140	1.260	1.160	0.696	1.351	1.081
Sharpe Ratio	0.576	0.650	0.404	0.555	0.649	0.625	0.638	0.368	0.613	0.632
Skewness	-0.354	-0.081	0.151	-0.464	-0.431	-0.752	-0.313	0.301	-0.167	-0.505
Kurtosis	5.683	5.340	3.336	5.736	3.969	6.118	4.639	4.678	5.748	4.296
Max Drawdown	-74.67	-61.71	-58.40	-71.28	-57.78	-74.06	-60.45	-69.03	-60.95	-68.26
Information Ratio	0.480	0.734	0.694	0.611	0.500	0.418	0.593	0.493	0.450	0.577

The panel A in the table documents that the mispricing strategy is more profitable than the whole market in all the regions. The long-only component is responsible for most of returns on the mispricing strategy. The short-only component then mainly serves as a hedge that increases Sharpe ratio and lowers maximum drawdown. The returns on long-only component of gradient

Table V Continued

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Panel B: Short-only Component of the Mispricing Strategy										
Weighted Least Squares										
Mean	0.297	0.386	-0.223	0.004	0.136	0.313	0.342	0.041	0.290	0.173
Sharpe Ratio	0.131	0.206	-0.112	0.002	0.077	0.166	0.198	0.021	0.118	0.109
Skewness	0.119	-0.557	0.376	-0.204	-0.234	-0.142	-0.744	0.565	0.045	-0.492
Kurtosis	5.883	4.473	3.564	4.335	4.644	5.892	4.741	3.966	5.797	5.035
Max Drawdown	-83.94	-85.78	-71.55	-85.03	-72.04	-78.71	-79.76	-81.47	-86.74	-67.48
Penalized Weighted Least Squares										
Mean	0.280	0.335	-0.243	-0.006	0.114	0.222	0.270	0.017	0.303	0.111
Sharpe Ratio	0.123	0.172	-0.121	-0.003	0.063	0.116	0.149	0.009	0.123	0.068
Skewness	0.123	-0.414	0.349	-0.070	-0.166	-0.024	-0.781	0.518	0.156	-0.510
Kurtosis	5.582	5.149	3.510	4.296	4.659	5.650	5.428	3.824	5.814	5.204
Max Drawdown	-83.56	-86.54	-70.06	-85.41	-72.34	-72.79	-78.49	-79.48	-86.34	-66.80
Gradient Boosting Regression Trees										
Mean	0.069	0.284	-0.497	-0.236	-0.085	-0.023	0.395	-0.358	-0.019	0.051
Sharpe Ratio	0.029	0.117	-0.219	-0.087	-0.042	-0.012	0.174	-0.162	-0.007	0.028
Skewness	-0.296	0.008	0.182	0.474	-0.224	-0.395	0.177	0.087	0.019	-0.353
Kurtosis	4.847	6.517	3.632	6.785	5.078	5.682	6.697	3.818	5.182	6.167
Max Drawdown	-79.96	-88.78	-66.42	-79.12	-68.60	-79.31	-87.79	-68.29	-78.39	-68.80
Random Forest										
Mean	0.077	0.237	-0.529	0.028	-0.080	-0.026	0.628	-0.408	-0.180	0.070
Sharpe Ratio	0.031	0.098	-0.237	0.010	-0.039	-0.012	0.274	-0.186	-0.067	0.037
Skewness	-0.263	0.185	0.288	0.723	-0.185	-0.294	0.404	0.173	0.755	-0.275
Kurtosis	4.697	7.201	3.800	8.629	4.968	5.249	9.564	3.687	9.069	6.330
Max Drawdown	-80.78	-90.00	-62.61	-90.70	-71.49	-79.21	-93.16	-65.70	-83.39	-73.58
Neural Networks										
Mean	-0.121	0.165	-0.539	-0.371	-0.206	-0.159	0.333	-0.404	0.175	-0.012
Sharpe Ratio	-0.054	0.075	-0.254	-0.142	-0.109	-0.084	0.164	-0.198	0.073	-0.007
Skewness	-0.353	-0.134	0.218	0.031	-0.210	-0.416	-0.184	0.226	-0.079	-0.370
Kurtosis	5.104	6.669	3.433	4.608	4.942	5.608	6.763	3.867	4.007	5.798
Max Drawdown	-80.06	-83.69	-67.02	-80.60	-67.65	-77.82	-80.68	-68.21	-86.50	-64.90

boosting regression tree mispricing strategy are about 5% a year larger than returns on the market. The other machine learning methods also outperform the market.

The more advanced machine learning methods outperform simple least squares both on the short side and long side. To conclude, the out-perfromance of the mispricing strategy is robust to short-selling constrains. Even short-selling-constrained investors can therefore benefit from the strategy.

B.2. Risk-adjusted Performance of the Strategy

Table VI
Performance of the Mispricing Strategy on Risk-adjusted Basis

The table shows returns of the mispricing strategy described in Table IV that is estimated on the individual stocks from the US adjusted for capital asset pricing model (CAPM) model and five Fama-French factors (FF5). The standard errors in t-statistics are adjusted for heterockedasticity and autocorrelation with Newey-West adjustment for up to 12 lags.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
					Weighted Least Squares					
Mean Return	0.801	0.680	0.922	0.782	0.810	0.575	0.647	0.648	0.633	0.639
	2.043	2.125	2.597	2.533	2.942	1.495	2.051	1.832	1.817	2.203
CAPM Alpha	0.951	0.672	0.930	0.744	0.851	0.688	0.588	0.657	0.707	0.666
	2.604	2.392	2.997	1.927	3.482	1.974	1.967	1.936	1.695	2.710
FF5 Alpha	0.328	0.133	0.672	0.232	0.263	0.080	0.134	0.402	-0.158	0.173
	1.114	0.543	2.189	0.562	1.336	0.262	0.486	1.121	-0.343	0.732
					Penalized Weighted Least Squares					
Mean Return	0.756	0.728	0.886	0.854	0.800	0.644	0.794	0.596	0.754	0.683
	1.960	2.191	2.538	2.451	2.800	1.599	2.632	1.645	1.810	2.294
CAPM Alpha	0.922	0.760	0.894	0.839	0.864	0.767	0.776	0.606	0.832	0.737
	2.505	2.629	2.844	2.032	3.391	2.132	2.513	1.750	1.825	2.869
FF5 Alpha	0.332	0.179	0.604	0.283	0.252	0.128	0.299	0.320	-0.079	0.192
	1.093	0.750	1.998	0.639	1.187	0.412	1.105	0.898	-0.160	0.836
					Gradient Boosting Regression Trees					
Mean Return	1.165	0.870	1.173	1.650	1.163	1.391	0.591	1.011	1.415	1.033
	4.266	2.978	4.021	5.470	6.641	4.465	2.198	2.662	5.236	4.998
CAPM Alpha	1.221	1.047	1.184	1.737	1.242	1.406	0.733	1.028	1.511	1.095
	3.630	3.589	3.643	4.883	5.999	3.726	2.539	2.408	4.109	4.283
FF5 Alpha	1.143	0.338	1.073	1.580	1.057	1.600	0.212	0.936	1.160	1.164
	3.680	1.116	3.222	3.841	4.377	4.834	0.730	2.211	2.595	4.535
					Random Forest					
Mean Return	1.050	1.037	1.107	0.943	1.074	0.977	0.339	1.028	1.183	0.798
	4.103	4.108	4.038	3.269	5.894	3.932	1.270	2.386	3.222	4.391
CAPM Alpha	1.157	1.200	1.118	1.035	1.172	1.050	0.475	1.039	1.318	0.887
	3.614	4.407	3.469	3.064	5.526	3.253	1.564	2.430	3.566	3.762
FF5 Alpha	0.782	0.433	0.919	0.871	0.727	0.740	-0.092	0.836	0.971	0.626
	2.748	1.502	2.787	2.258	3.124	2.626	-0.275	1.974	2.304	2.343
					Neural Networks					
Mean Return	1.416	1.097	1.295	1.752	1.346	1.420	0.826	1.100	1.177	1.093
	4.336	3.734	5.759	4.917	7.829	4.442	3.352	3.240	3.342	5.627
CAPM Alpha	1.402	1.179	1.301	1.788	1.383	1.354	0.903	1.108	1.247	1.103
	3.928	4.057	5.257	4.575	7.853	3.471	3.413	3.285	3.540	4.586
FF5 Alpha	1.482	1.038	1.185	1.435	1.334	1.584	0.758	1.008	0.749	1.323
	5.018	3.885	4.581	3.527	7.754	5.009	2.985	2.941	2.257	5.895

We have so far focused only on raw returns on the mispricing strategy without accounting for any risk factors. Table VI presents performance of the strategy after accounting for market returns and five Fama-French factors. Accounting for market return should have little impact on the performance of the strategy since it is long-short, and thus close to market neutral, by construction. Table VI confirms that it is indeed the case and capital asset pricing model (CAPM) alpha is close to the mean returns for all the estimation methods. The results are, however, very different when adjusting for five Fama-French factors. There is again almost no difference between the mean returns and alphas for more complicated estimation methods but there is a visible deterioration in risk-adjusted performance for the linear estimation methods. The linear estimation methods therefore lead to mispricing signal that is close to the traditional risk factors.

To conclude, the profitability of the mispricing strategy is significant even at risk-adjusted basis. The more complicated estimation methods then lead to returns that are unrelated to the traditional risk factors.

B.3. Variable Importance

One of the disadvantages of more complex machine learning methods is the difficulty in interpreting the resulting models because of potentially high-dimensional and nonlinear interactions among variables. The main goal of our study is superior out-of-sample performance even at the cost of inability to fully interpret all the interactions of variables in the resulting models. Limited interpretation of the fitted models will now follow.

That being said, inspired by Chen, Pelger, and Zhu (2019), Sirignano, Sadhwani, and Giesecke (2016), and Horel and Giesecke (2019) we now examine importance of individual variables. Variable importance VI_j for variable j is defined in equation 4 as an elasticity of predicted (region-wise demeaned) returns to changes in the individual characteristics used as predictors.

$$VI_j = \sum_{t \in T} \sum_{i \in N_t} \frac{\partial \hat{r}_{it}(x_{i,t-1,1}, x_{i,t-1,2}, ..., x_{i,t-1,M})}{\partial x_j} \tag{4}$$

The variable importance is calculated for each characteristic in various settings. Figures 7 and 8 show the variable importance across regions for the fifty most important variables around the Globe on all-but-micro-caps and liquid-only universes of stocks for the mispricing strategy as described in Table IV. We document substantial differences in the rank of variable importance between liquid and all-but-micro-caps stocks in all regions, where for example "Bid-Ask Spread" falls from being the most important global variable on all-but-micro-caps stocks to the second least important global variable on the liquid stocks.[25] The Spearman's rank correlation coefficient between variable importance ranks for all-but-micro-caps and liquid universe of stocks is only 0.479.

Table VII shows the Spearman's rank correlation of variable importance scores across the

[25] The least important variables are not visible in the Figure 8.

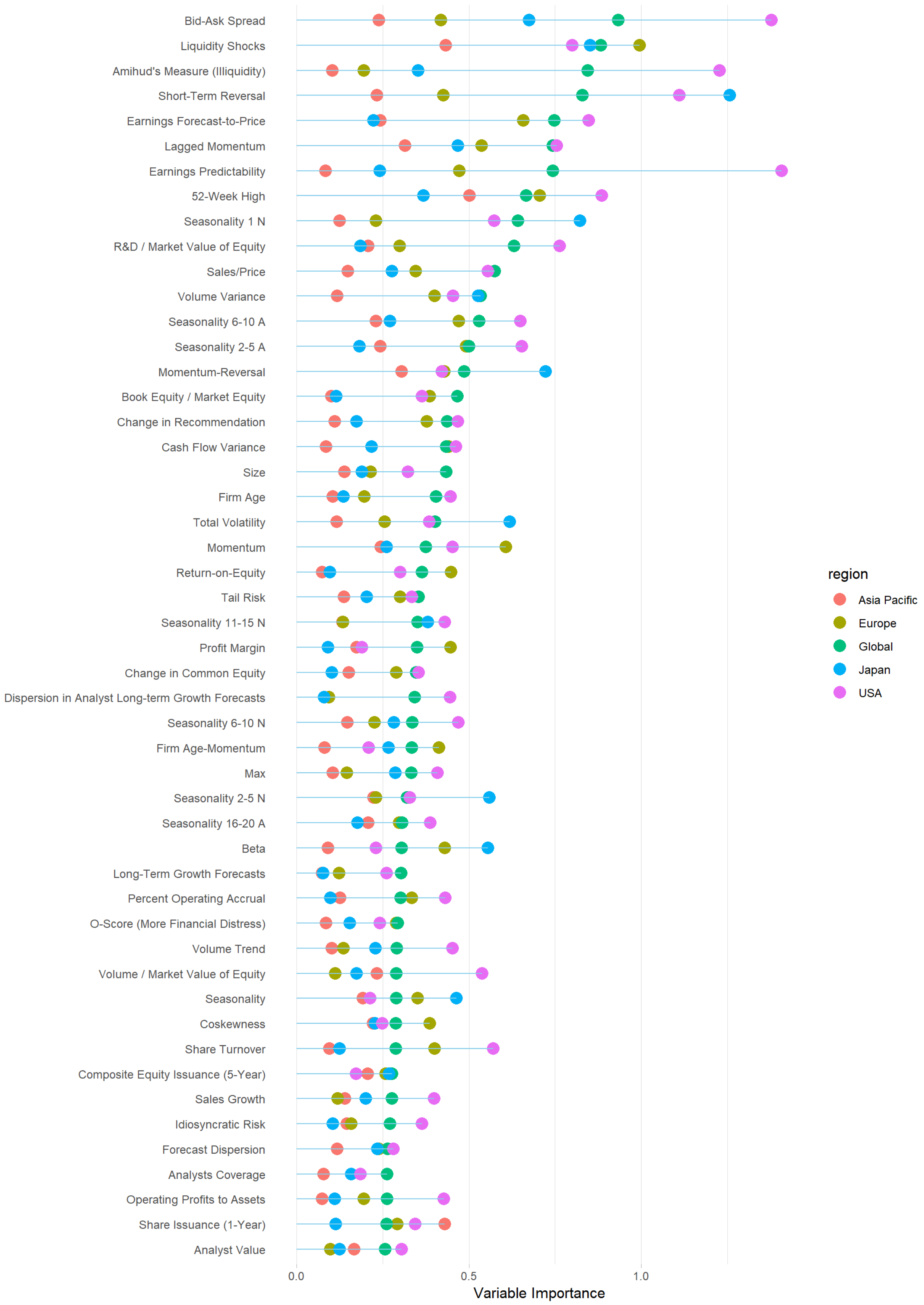


Figure 7. Variable Importance on All-but-micro-caps Universe of Stocks. The figure shows variable importance as defined in Equation 4 for the 50 globally most important variables for the mispricing strategy as described in Table IV that is estimated on the individual stocks from the US.

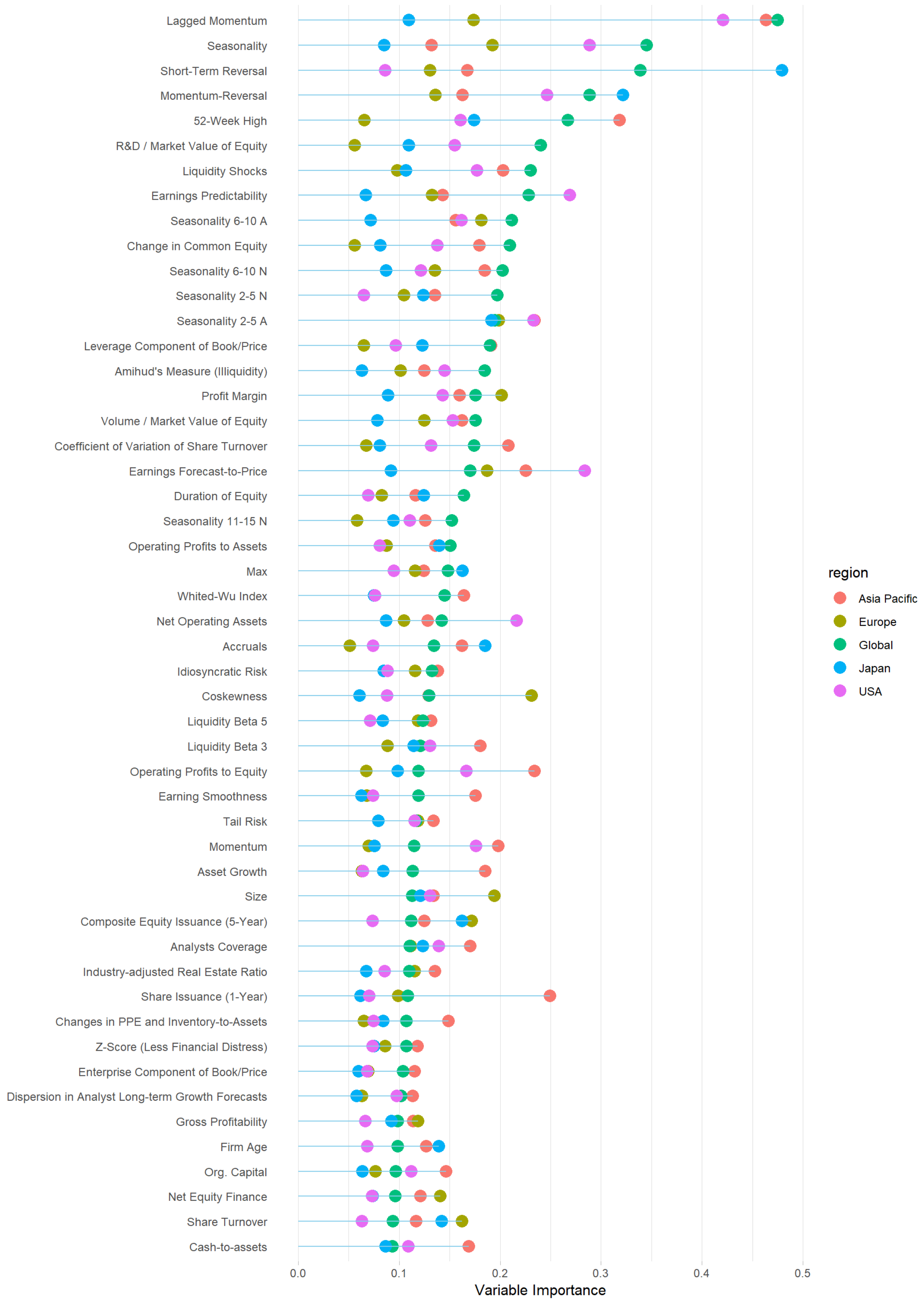


Figure 8. Variable Importance on Liquid Universe of Stock. The figure shows variable importance as defined in Equation 4 for the 50 globally most important variables for the mispricing strategy as described in Table IV that is estimated on the individual stocks from the US.

regions under various forecasting methods. There is a great heterogeneity in the ranks of variable importance across the regions. This sheds some light on the limited value of international evidence as will be documented in Section III. More importantly, the predictions from the US perform as well as predictions from the individual respective regions despite having only loosely connected variable importance.

Table VII
Spearman's Correlation Matrices for Region-specific Variable Importance

The table shows Spearman correlation matrices between region-specific ranks of variable importance for mispricing strategy using OLS, gradient boosting regression trees and neural networks as described in Section I.D.

	OLS					GBRT					NN				
	Global	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP
Global	1.000	0.552	0.610	0.564	0.576	1.000	0.745	0.772	0.586	0.754	1.000	0.739	0.672	0.605	0.325
USA	0.552	1.000	0.406	0.363	0.397	0.745	1.000	0.706	0.541	0.591	0.739	1.000	0.531	0.460	0.341
Europe	0.610	0.406	1.000	0.482	0.443	0.772	0.706	1.000	0.606	0.631	0.672	0.531	1.000	0.478	0.318
Japan	0.564	0.363	0.482	1.000	0.541	0.586	0.541	0.606	1.000	0.507	0.605	0.460	0.478	1.000	0.404
AP	0.576	0.397	0.443	0.541	1.000	0.754	0.591	0.631	0.507	1.000	0.352	0.341	0.318	0.404	1.000

Table VIII shows differences in the ranks of variable importance for mispricing strategy under different forecasting methods. The Spearman's rank correlation coefficient between neural networks and gradient boosting regression trees is only 0.417 while both methods deliver significant out-of-sample profitability when used as forecasting methods in the mispricing strategy.

Table VIII
Spearman's Correlation for Method-specific Variable Importance

	OLS	GBRT	NN
OLS	1.000	0.384	0.454
GBRT	0.384	1.000	0.417
NN	0.454	0.417	1.000

III. The Role of International Evidence

The evidence so far documented that the anomalies identified, based on the past data in the US, are profitable out-of-sample in all the regions. Can international data outside the US be used to better select the winning strategies?

There are some arguments for the usefulness of the international data. The international data increases sample size and therefore limits the possibility for data-mining and in-sample overfitting. The larger sample size also generally provides larger power to statistical tests which should lead to more precise selection of truly significant strategies. One crucial requirement for the tangible benefit of the new observations is that they are independent from the original observations. Table II has documented that portfolio returns on the individual anomalies are only mildly correlated over time across the regions, which suggests that the international observations are independent to a large extent. The international evidence extends the sample size mainly in the most recent period.

The most recent data is also the most useful as the financial markets are changing rapidly and the older data may not be relevant anymore.

There are, however, also some problems with suitability of the international evidence. The individual global regions have very different institutional settings. Bankruptcy laws, tax laws, investor protection, and accounting standards vary widely across the regions. The institutional differences can lower the usefulness of historical data outside the respective regions. The larger estimation sample improves forecasts through consistency. The consistency, however, works only if the underlying true drivers of stock returns are uniform over the regions, which is in no way guaranteed.

A. *Profitability of the Individual Anomalies*

The role of international evidence is first studied on the individual anomalies. The individual anomalies are suitable to assess the benefit of observing past performance of individual quantitative strategies outside the US for selection of strategies that are the most profitable out-of-sample.

Table IX provides regressions of future Sharpe ratios of portfolio returns from the individual anomalies regressed on their past Sharpe ratios in Europe (E), the USA, and Japan (J). The analysis is restricted to these three regions as there are historically only few liquid stocks in the Asia Pacific region. The liquid universe of stocks used elsewhere in this study is too restrictive here and many implications could remain hidden. We therefore also present results for a more liquid sample of stocks. The long-short portfolios are created from sample of stocks restricted to those with market cap larger than the lowest decile on NYSE (All-but-micro-caps) or from a more liquid sample with stricter requirement on both capitalization and dollar traded volume (Liquid Universe).[26]

Panel A of the table uses ordinary least squares regressions of Sharpe ratios from the original estimation sample period in anomalies' publications regressed on post-publication five-year Sharpe ratios in the three regions. The post-publication Sharpe ratios are used because the returns are very noisy and do not lead to any significant finding. The methodology is suited to answer the question whether adding international evidence to the original papers could have decreased the data-mining and post-publication decay.

It is apparent that evidence from Japan and Europe is not useful for predictions of post-publication performance in the US. The past performance in the US is, however, significantly useful for predictions of performance inside and outside of the US at 5% level for equal-weighted portfolios on all-but-micro-caps sample of stocks. The estimated coefficients imply that about 30% of the in-sample Sharpe ratio in the US persists in all the regions for five years after the anomalies

[26]See Section I.A for more precise definitions of the samples.

Table IX
Predictive Power of the Past Performance of Anomalies in Different Regions

Panel A presents results from regressions of 5-year post-publication Sharpe ratios of the individual anomalies on in-sample Sharpe ratios from Japan (J), Europe (E), and the USA. The in-sample period is the same as in the studies where the individual anomalies were published. Panel B then shows regressions of future three-year (out-of-sample) Sharpe ratios on past (in-sample) Sharpe ratios of the individual anomalies. The regressions are based on a panel of Sharpe ratios where the future ratios are estimated over non-overlapping three-year intervals. The first three-year interval starts at the end of June 1995 and the last interval ends in June 2016. The in-sample Sharpe ratios are based on the longest possible estimation sample preceding the three-year out-of-sample period. Only anomalies that have been published by the start of the given out-of-sample period are included in the panel. The portfolios are created from sample of stocks restricted to those with market cap larger than the lowest decile on NYSE (All-but-micro-caps) or from a more liquid sample with stricter requirement on both capitalization and dollar traded volume (Liquid Universe). The standard errors are adjusted for heteroskedasticity as in Driscoll and Kraay (1998).

	Dependent Variables (Sharpe Ratios) are From											
	USA				Europe				Japan			
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Panel A: 5-year Post-publication Sharpe Ratios Regressed on In-sample Sharpe Ratios												
Equal-weighted Portfolios from the All-but-micro-caps												
Intercept	0.07	0.17	0.15	0.08	0.22	0.42	0.43	0.25	0.05	0.17	0.20	0.03
	(1.29)	(3.86)	(4.01)	(1.39)	(3.20)	(6.69)	(7.99)	(2.90)	(0.85)	(3.05)	(4.29)	(0.43)
In-sample SR USA	0.21			0.32	0.37			0.56	0.32			0.48
	(2.24)			(2.53)	(3.33)			(3.44)	(3.18)			(3.78)
In-sample SR E		-0.07		-0.18		0.05		-0.12		0.08		-0.09
		(-1.12)		(-2.68)		(0.64)		(-1.57)		(1.23)		(-1.28)
In-sample SR J			0.06	0.07			0.10	0.07			0.18	0.14
			(0.90)	(1.15)			(1.22)	(0.81)			(2.92)	(2.19)
Sample Size	139	109	108	108	139	109	108	108	139	109	108	108
R-Squared	0.04	0.01	0.01	0.08	0.06	0.00	0.01	0.09	0.06	0.01	0.04	0.12
Value-weighted Portfolios from the All-but-micro-caps												
Intercept	0.10	0.14	0.11	0.14	0.14	0.24	0.23	0.19	0.07	0.12	0.14	0.05
	(1.77)	(2.90)	(3.01)	(2.03)	(2.58)	(5.22)	(5.53)	(2.79)	(1.35)	(2.81)	(3.39)	(0.73)
In-sample SR USA	0.02			0.01	0.20			0.22	0.20			0.32
	(0.16)			(0.05)	(1.45)			(1.30)	(1.48)			(1.72)
In-sample SR E		-0.11		-0.12		-0.06		-0.10		0.05		0.00
		(-1.53)		(-1.66)		(-0.84)		(-1.43)		(0.70)		(0.01)
In-sample SR J			-0.03	0.01			0.04	0.06			0.04	0.03
			(-0.39)	(0.12)			(0.35)	(0.57)			(0.59)	(0.45)
Sample Size	139	109	108	108	139	109	108	108	139	109	108	108
R-Squared	0.00	0.02	0.00	0.02	0.02	0.00	0.00	0.03	0.02	0.00	0.00	0.04
Equal-weighted Portfolios from the Liquid Universe												
Intercept	0.04	0.08	0.08	0.06	0.13	0.19	0.20	0.17	0.01	0.07	0.10	0.00
	(0.98)	(1.85)	(2.32)	(0.80)	(3.27)	(4.73)	(5.37)	(2.83)	(0.23)	(1.46)	(2.27)	(0.05)
In-sample SR USA	0.05			0.15	0.15			0.17	0.22			0.35
	(0.49)			(0.80)	(1.16)			(0.71)	(1.53)			(1.62)
In-sample SR E		-0.05		-0.06		0.02		-0.00		0.08		0.03
		(-0.75)		(-0.96)		(0.55)		(-0.08)		(1.21)		(0.49)
In-sample SR J			-0.01	-0.01			0.08	0.07			0.12	0.09
			(-0.21)	(-0.12)			(1.27)	(1.09)			(1.94)	(1.27)
Sample Size	139	109	107	107	139	109	107	107	137	107	105	105
R-Squared	0.00	0.01	0.00	0.01	0.01	0.00	0.01	0.02	0.02	0.01	0.02	0.05
Value-weighted Portfolios from the Liquid Universe												
Intercept	0.03	0.05	0.05	0.05	0.13	0.15	0.14	0.16	0.02	0.08	0.09	-0.01
	(0.81)	(1.29)	(1.23)	(0.92)	(3.76)	(4.61)	(4.68)	(3.74)	(0.38)	(2.06)	(2.38)	(-0.11)
In-sample SR USA	0.03			0.04	-0.03			-0.02	0.24			0.45
	(0.30)			(0.23)	(-0.27)			(-0.15)	(1.54)			(2.19)
In-sample SR E		-0.09		-0.09		-0.06		-0.07		0.02		-0.01
		(-1.55)		(-1.49)		(-1.04)		(-1.26)		(0.33)		(-0.14)
In-sample SR J			-0.03	-0.01			0.05	0.06			0.12	0.09
			(-0.54)	(-0.18)			(0.60)	(0.82)			(1.25)	(1.06)
Sample Size	139	109	107	107	139	109	107	107	137	107	105	105
R-Squared	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.02	0.00	0.01	0.07

are published. The predictability of the regressions is much higher for equal-weighted portfolios on all-but-micro-caps sample of stocks relative to value-weighting or liquid universe of stocks where

Table IX Continued

	Dependent Variables (Sharpe Ratios) are From											
	USA				Europe				Japan			
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Panel B: Future 3-year Sharpe Ratios Regressed on Past Sharpe Ratios												
Value-weighted Portfolios from the All-but-micro-caps												
Intercept	-0.02	0.14	0.21	-0.02	0.16	0.33	0.56	0.16	0.11	0.24	0.24	0.11
	(-0.34)	(3.05)	(5.40)	(-0.31)	(2.31)	(5.85)	(11.10)	(2.32)	(1.77)	(4.69)	(5.54)	(1.71)
In-sample SR USA	0.74			0.80	1.24			0.93	0.52			0.56
	(3.96)			(4.36)	(10.20)			(4.32)	(4.55)			(2.94)
In-sample SR E		0.23		-0.04		0.66		0.36		0.12		-0.11
		(2.89)		(-0.44)		(4.44)		(2.33)		(1.79)		(-0.73)
In-sample SR J			0.08	-0.05			0.06	-0.17			0.28	0.21
			(0.88)	(-0.57)			(0.29)	(-1.37)			(2.17)	(1.68)
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sample Size	530	530	530	530	530	530	530	530	527	527	527	527
R-Squared	0.06	0.02	0.00	0.06	0.10	0.08	0.00	0.12	0.02	0.00	0.02	0.03
Value-weighted Portfolios from the All-but-micro-cap												
Intercept	0.02	0.09	0.15	0.01	0.17	0.27	0.32	0.17	0.11	0.18	0.20	0.11
	(0.33)	(2.19)	(4.12)	(0.20)	(3.19)	(6.24)	(8.40)	(3.10)	(1.94)	(4.04)	(5.02)	(1.95)
In-sample SR USA	0.58			0.48	0.63			0.57	0.37			0.39
	(2.42)			(2.00)	(2.77)			(3.86)	(3.75)			(5.19)
In-sample SR E		0.26		0.17		0.21		0.09		0.07		-0.02
		(2.31)		(1.60)		(1.11)		(0.53)		(0.76)		(-0.17)
In-sample SR J			-0.03	-0.12			0.00	-0.07			0.01	-0.02
			(-0.67)	(-2.22)			(0.03)	(-1.06)			(0.09)	(-0.27)
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sample Size	530	530	530	530	530	530	530	530	527	527	527	527
R-Squared	0.02	0.01	0.00	0.03	0.02	0.01	0.00	0.02	0.01	0.00	0.00	0.01
Equal-weighted Portfolios from the Liquid Universe												
Intercept	0.01	0.08	0.11	-0.00	0.14	0.18	0.22	0.13	0.01	0.04	0.09	0.00
	(0.17)	(5.67)	(3.52)	(-0.20)	(3.79)	(6.97)	(5.59)	(3.87)	(0.74)	(3.45)	(2.88)	(0.09)
In-sample SR USA	0.64			0.60	0.48			0.40	0.53			0.38
	(3.65)			(2.70)	(2.19)			(1.57)	(6.26)			(4.11)
In-sample SR E		0.16		0.09		0.19		0.10		0.28		0.13
		(2.15)		(0.62)		(1.47)		(0.74)		(4.41)		(1.29)
In-sample SR J			0.04	-0.03			0.16	0.09			0.30	0.22
			(1.18)	(-0.42)			(1.27)	(0.96)			(2.99)	(1.94)
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sample Size	523	523	523	523	522	522	522	522	515	515	515	515
R-Squared	0.03	0.01	0.00	0.03	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.03
Value-weighted Portfolios from the Liquid Universe												
Intercept	0.01	0.05	0.07	0.00	0.16	0.16	0.18	0.15	0.05	0.09	0.13	0.03
	(0.21)	(2.57)	(3.11)	(0.03)	(3.81)	(5.18)	(3.55)	(3.27)	(2.22)	(7.65)	(8.12)	(0.99)
In-sample SR USA	0.37			0.35	0.11			0.06	0.53			0.42
	(1.55)			(1.20)	(0.38)			(0.21)	(3.77)			(3.21)
In-sample SR E		0.10		0.09		0.13		0.10		0.32		0.30
		(0.68)		(0.43)		(0.61)		(0.49)		(3.84)		(2.82)
In-sample SR J			-0.03	-0.07			0.09	0.06			0.02	-0.09
			(-1.10)	(-1.54)			(0.82)	(1.04)			(0.23)	(-1.05)
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sample Size	523	523	523	523	522	522	522	522	515	515	515	515
R-Squared	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.02

almost none of the coefficients are significant at 5% level.

Panel B studies the predictability of future Sharpe ratios in panel setting. Past Sharpe ratios in the three regions and future three-year Sharpe ratios are estimated at the end of June every three years starting in 1995.[27] Only anomalies that were published by the given June are retained in the sample. All the regressions use fixed effects for time periods to control for time variability of profitability of the anomalies. The panel setting is suitable to answer the question whether past international performance is useful in estimation of future performance of the already published anomalies and thus to help to pick winning strategies at any point in time.

The table shows some predictability of the future performance using past performance in the respective regions. The past performance in the US is, however, a dominant predictor when the past performance from all the regions is included together. The predictability is generally higher for equal-weighted portfolios and all-but-micro-caps sample of stocks than for value-weighted portfolios and liquid universe of stocks.

The puzzling result that past performance of the anomalies outside the US is not useful for predictions of their future returns in the US can be explained by the fact that strategies that work everywhere should attract the largest attention from investors. The investors should in turn drive their future returns down.

B. Portfolio-mixing Strategy

Table X studies the benefits of past international evidence for selection of anomalies with better out-of-sample performance in the portfolio-mixing strategy. It shows performance of portfolios created by equally combining individual portfolios for all the signals significant with 5% false discovery rate. The performance metrics are based on returns on portfolios over the July 1995 to December 2016 period. The significance is determined by the past p-values on intercept in panel regressions of portfolio returns on just intercept. The p-values are based on heteroskedasticity and autocorrelation robust standard errors as in Driscoll and Kraay (1998). The portfolio returns are from the US, the US & Japan, the US & Europe, or the US & Japan & Europe. The standard errors used for computation of p-values are HAC robust. The significant signals are chosen at the end of each June and the past p-values are taken for the period from July 1963 (1990) for the US (Japan and Europe) up to the time of portfolio formation.

The table X supports the evidence from table IX. The addition of international evidence does not lead to any significant improvement in out-of-sample performance of the strategy that mixes the significant anomalies.

[27]The year 1995 was chosen so that there is enough evidence in Europe and Japan. The choice of starting point does not affect the conclusions.

Table X
Does International Evidence Improve the Performance of the Portfolio-mixing Strategy?

The table shows returns on the portfolio-mixing strategy that equally invests in portfolios of the individual significant anomalies as described in Section I.C.2. The significance of anomalies is based on mean portfolio returns in the US, the US with Japan; the US with Europe; or the US with Japan plus Europe.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Evidence from the US										
Mean	0.174	0.297	0.001	0.663	0.284	0.301	0.180	0.253	0.882	0.404
Sharpe Ratio	0.227	0.484	0.002	0.695	0.566	0.387	0.270	0.198	0.816	0.598
Skewness	0.083	-0.085	-1.885	-1.087	-0.436	0.356	0.197	-0.046	1.871	1.358
Kurtosis	9.963	9.230	14.68	13.16	6.920	6.481	9.243	25.09	16.59	22.41
Max Drawdown	-29.40	-17.96	-27.63	-27.43	-12.95	-18.12	-26.33	-61.07	-17.35	-20.79
Evidence from the US & Japan										
Mean	0.144	0.277	0.071	0.476	0.242	0.192	0.181	0.245	0.949	0.392
Sharpe Ratio	0.212	0.475	0.100	0.594	0.542	0.250	0.271	0.162	0.796	0.501
Skewness	0.672	0.026	-3.252	-0.929	0.282	0.458	0.624	0.198	0.885	1.681
Kurtosis	12.56	10.73	31.12	9.678	8.659	9.811	8.543	27.68	17.66	29.47
Max Drawdown	-28.32	-17.73	-37.96	-21.17	-15.93	-20.70	-25.12	-69.75	-23.55	-25.91
Information Ratio	-0.088	-0.080	0.191	-0.389	-0.174	-0.325	0.002	-0.018	0.118	-0.052
Evidence from the US & Europe										
Mean	0.150	0.334	0.191	0.502	0.294	-0.091	0.023	0.234	0.445	0.153
Sharpe Ratio	0.203	0.545	0.332	0.547	0.624	-0.124	0.037	0.197	0.269	0.245
Skewness	1.195	1.953	-0.842	-1.111	1.278	-0.325	0.491	0.497	1.991	1.275
Kurtosis	19.90	15.37	10.28	17.66	15.70	7.029	6.516	21.84	20.58	11.89
Max Drawdown	-33.19	-14.39	-20.88	-30.54	-18.08	-38.86	-37.71	-52.86	-45.93	-21.51
Information Ratio	-0.052	0.100	0.450	-0.327	0.038	-0.439	-0.232	-0.019	-0.270	-0.378
Evidence from the US & Japan & Europe										
Mean	-0.001	0.182	0.098	0.346	0.156	-0.088	0.067	0.154	0.440	0.143
Sharpe Ratio	-0.002	0.440	0.145	0.455	0.422	-0.144	0.113	0.136	0.342	0.282
Skewness	-1.760	0.337	-2.962	-1.259	-0.877	-0.223	0.198	-0.270	-0.202	0.150
Kurtosis	12.25	5.283	29.25	16.60	7.187	8.267	7.281	20.28	13.43	8.314
Max Drawdown	-33.36	-13.59	-35.17	-23.61	-15.07	-33.05	-30.14	-62.00	-48.76	-12.54
Information Ratio	-0.238	-0.236	0.192	-0.456	-0.289	-0.460	-0.202	-0.113	-0.326	-0.461

C. Mispricing Strategy

The portfolio level analysis of individual anomalies found little value for the international evidence. The question is revisited here with the machine learning methods combining anomalies into one mispricing signal.

The previous machine learning evidence was based on predictive regressions estimated solely on data from the US. This section first investigates whether estimating the predictive regressions in the respective regions is more suitable than estimating them only on data from the US. It then explores whether combining estimation samples from the individual regions can improve the profitability to the mispricing strategy.

There is surprisingly only a small difference between returns on strategies that are estimated on data from the US in table IV and those that are estimated on data in the respective regions in table XI. One explanation for the similarity is that the sample size in the US is already large enough to capture the true drivers of stock returns that are globally valid. One exception is Asia

Pacific region where there are only a few liquid stocks historically, which makes the predictive regressions imprecise. The performance of the mispricing strategy in Japan is also notably worse than when estimated on the US data. The explanation is again simple. Japan has undergone a slow eruption of an asset price bubble at the beginning of the estimation sample in early 1990s. The estimated relationships that are valid for this specific period fare badly out-of-sample where the stock market dynamics go back to their normal state.

Table XI
Performance of the Mispricing Strategy Estimated in the Individual Regions

The table shows out-of-sample performance of the mispricing strategy as described in Table IV. The predictive regressions for individual stock returns are estimated in each respective region.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Weighted Least Squares										
Mean	0.801	0.682	0.750	1.117	0.811	0.575	0.483	0.348	0.876	0.596
Sharpe Ratio	0.479	0.450	0.396	0.507	0.786	0.348	0.338	0.157	0.375	0.541
Skewness	-0.340	-1.280	0.353	-1.957	-0.200	-0.121	-0.995	-0.255	-1.500	-0.349
Kurtosis	8.521	10.03	7.663	20.65	6.543	7.214	7.863	7.841	18.58	6.483
Max Drawdown	-64.70	-50.52	-66.56	-61.71	-36.44	-69.75	-42.84	-72.33	-63.57	-44.58
Penalized Weighted Least Squares										
Mean	0.756	0.753	0.701	1.333	0.823	0.644	0.616	0.423	1.359	0.656
Sharpe Ratio	0.443	0.458	0.363	0.535	0.749	0.381	0.393	0.197	0.534	0.577
Skewness	-0.487	-1.026	0.360	-1.665	-0.278	-0.316	-1.033	0.329	-1.072	-0.376
Kurtosis	8.703	8.700	7.280	21.06	6.723	7.297	8.447	6.321	15.03	7.534
Max Drawdown	-65.36	-46.61	-66.88	-69.25	-39.84	-68.02	-43.74	-58.29	-63.68	-39.76
Gradient Boosting Regression Trees										
Mean	1.165	0.725	0.951	1.766	1.107	1.391	0.319	0.678	1.522	0.915
Sharpe Ratio	0.720	0.596	0.636	0.761	1.183	0.831	0.238	0.400	0.581	0.850
Skewness	0.319	-0.884	0.445	-0.346	-0.012	0.561	-1.250	0.071	0.026	-0.450
Kurtosis	6.653	7.508	5.686	19.50	5.987	9.287	7.699	4.559	13.42	7.646
Max Drawdown	-38.31	-45.14	-34.11	-56.04	-22.56	-43.93	-58.13	-55.73	-55.96	-31.15
Random Forest										
Mean	1.050	0.353	1.022	0.960	0.892	0.977	0.140	0.792	1.112	0.711
Sharpe Ratio	0.703	0.265	0.779	0.544	1.007	0.691	0.094	0.503	0.516	0.688
Skewness	-0.328	-1.281	-0.201	0.591	-0.408	-0.594	-1.111	0.201	0.768	-0.953
Kurtosis	5.989	9.421	4.537	6.862	6.323	4.951	6.857	4.150	9.382	6.801
Max Drawdown	-30.69	-51.84	-32.79	-52.27	-22.31	-30.59	-60.13	-47.42	-51.77	-29.88
Neural Networks										
Mean	1.416	0.748	0.958	1.192	1.133	1.420	0.561	0.616	0.986	0.988
Sharpe Ratio	0.905	0.544	0.572	0.592	1.308	0.905	0.383	0.305	0.423	1.025
Skewness	-0.083	-0.637	0.464	-0.435	0.026	-0.167	-0.696	0.054	-0.151	-0.465
Kurtosis	7.316	6.991	7.300	5.796	4.891	6.432	7.182	8.016	5.195	6.920
Max Drawdown	-44.60	-50.30	-48.09	-55.31	-18.16	-48.11	-37.60	-72.91	-68.84	-21.88

Table XII shows mean returns and other performance statistics for gradient boosting regression trees mispricing strategy as in table IV. The only difference with respect to table IV is that the future individual stock returns are predicted from regressions estimated on historical data that are not solely from the US. Predictive regressions with training sample from the US, the US & Japan, the US & Europe, or the US & Japan & Europe & Asia Pacific are compared. These three regions cover most of the developed markets and global stock market capitalization. Corresponding evidence for least square mispricing strategy is provided in the online appendix.[28]

[28]It is omitted here for the sake of space as all the findings are very similar to gradient boosting regression trees.

Table XII
Performance of the Mispricing Strategy Estimated on the International Data

The table shows returns of the mispricing strategy based on gradient boosting regression trees described in Table IV. The historical predictive regressions are estimated on individual stocks from combinations of the four covered regions: the US, Japan, Europe, Europe, Asia Pacific.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Estimated in the US										
Mean	1.165	0.870	1.173	1.650	1.163	1.391	0.591	1.011	1.415	1.033
Sharpe Ratio	0.720	0.644	0.766	1.005	1.146	0.831	0.412	0.525	0.800	0.870
Skewness	0.319	-1.160	0.682	-0.437	-0.449	0.561	-1.314	0.800	-0.112	-0.433
Kurtosis	6.653	10.17	8.274	5.575	6.812	9.287	12.03	8.611	4.718	7.797
Max Drawdown	-38.31	-48.25	-34.37	-36.65	-27.45	-43.93	-42.31	-41.79	-39.58	-35.62
Estimated in the US and Cross-validated on the International Data										
Mean	1.114	1.012	1.021	1.192	1.085	1.318	0.665	0.675	0.846	0.968
Sharpe Ratio	0.704	0.747	0.686	0.655	1.057	0.776	0.473	0.355	0.410	0.785
Skewness	0.215	-0.422	0.313	-1.584	-0.558	0.270	-0.673	0.474	-1.266	-0.146
Kurtosis	6.706	11.44	8.626	12.11	7.354	7.886	13.89	8.906	12.23	9.259
Max Drawdown	-36.49	-44.69	-39.01	-49.81	-30.07	-39.70	-43.57	-44.37	-61.01	-35.13
Information Ratio	-0.108	0.235	-0.285	-0.415	-0.265	-0.108	0.084	-0.399	-0.379	-0.115
Estimated in the US & Japan										
Mean	1.317	1.015	1.353	1.537	1.289	1.616	0.812	1.001	1.722	1.262
Sharpe Ratio	0.809	0.911	1.030	1.016	1.413	0.907	0.647	0.550	1.008	1.093
Skewness	0.602	0.715	0.150	0.176	0.380	0.876	0.088	0.837	0.362	0.828
Kurtosis	7.612	8.349	6.909	3.599	8.273	10.26	9.448	10.76	3.516	9.092
Max Drawdown	-34.18	-25.91	-25.82	-29.65	-19.15	-42.98	-30.11	-45.32	-29.95	-31.87
Information Ratio	0.186	0.153	0.172	-0.083	0.224	0.208	0.203	-0.007	0.191	0.295
Estimated in the US & Europe										
Mean	1.361	1.016	1.173	1.555	1.268	1.513	0.716	0.786	1.397	1.111
Sharpe Ratio	0.854	0.812	0.763	0.892	1.241	0.875	0.501	0.431	0.654	0.944
Skewness	0.049	0.159	0.251	-1.266	-0.313	0.537	-0.463	-0.063	-1.763	-0.055
Kurtosis	6.672	6.745	6.599	11.62	6.270	8.410	7.574	5.445	17.04	6.332
Max Drawdown	-38.16	-40.92	-34.41	-45.58	-27.77	-38.96	-47.54	-49.58	-56.36	-30.43
Information Ratio	0.281	0.197	-0.000	-0.085	0.248	0.150	0.129	-0.214	-0.010	0.130
Estimated in the US & Japan & Europe & Asia Pacific										
Mean	1.325	1.009	1.281	2.295	1.373	1.432	0.955	1.066	2.317	1.225
Sharpe Ratio	0.808	0.870	0.960	1.486	1.394	0.803	0.680	0.615	1.056	1.048
Skewness	0.257	0.262	0.114	0.624	-0.034	0.991	0.745	-0.013	0.750	0.451
Kurtosis	7.133	5.071	5.931	5.360	7.274	11.46	7.285	4.745	8.252	7.276
Max Drawdown	-37.15	-33.19	-24.69	-20.31	-26.01	-38.84	-26.46	-44.00	-53.48	-30.43
Information Ratio	0.191	0.131	0.114	0.530	0.404	0.036	0.295	0.045	0.483	0.251

The table provides mixed results on the value of international evidence. There is a small gain from adding the international stocks to local training sample in the US for equal-weighted portfolios. Historical data in the US is therefore completely sufficient for the future predictions in the US. Profitability of the mispricing strategy in Europe improves with predictions based on the estimation sample from the US and Europe relative to from the US only. The profitability in Japan also improves with training sample from both the US and Japan instead of from the US only. The largest gains in profitability are in Asia Pacific region where training samples from Japan and Europe are jointly beneficial.

The table also shows the gradient boosting regression tree mispricing strategy estimated in the US using parameters cross-validated in the other three regions. The cross-validation on data outside the US could add some predictive power as the validation sample is coming from more recent period than when the training sample is from the US only. The table, however, documents that there is no gain from cross-validating outside the US.

To conclude, the regional institutional setting is indeed an important determinant of stock return drivers. There is no gain for the US investor to seek international evidence for quantitative strategies. The larger statistical power, caused by a larger sample, seems to be completely offset by the differences in institutional setting.

IV. Importance of New Anomalies for Profitability of the Strategies

Figure 1 documented that the number of published anomalies is increasing roughly linearly over time. Harvey et al. (2016) found even sharper increase for published as well as unpublished anomalies. Researchers are looking at the same data again and again to find the new anomalies which should lead to a large proportion of false positive discoveries. The proportion of false discoveries is expected to increase over time as the strongest anomalies are likely already published. Harvey et al. (2016) therefore concluded that most of the recently published studies can be explained by data-mining and the standard critical values for statistical significance no longer apply. The data-mining should also lead to a lower predictive power of the new anomalies. Individual studies introducing new anomalies almost never properly control for all anomalies published previously. Many of the new anomalies are therefore subsumed by existing anomalies in proper multiple hypothesis setting as documented by Green et al. (2017).

Most of the widely accepted anomalies have been published before 1995.[29] It is therefore worth studying whether the more recently published drivers of stock returns are also important. This section investigates the marginal value of recently published anomalies for profitability of the mispricing strategy after accounting for anomalies published earlier.

Table XIII presents mean returns and Sharpe ratios on the mispricing strategy as specified in table IV but with further restrictions on the universe of anomalies. The mispricing strategy is estimated using anomalies that were published before 1995, 2000, or 2005. Its performance is then tracked over the 2005-2016 period.[30] The different sets of anomalies provide a good indication for marginal value of the new signals published after 1995, while accounting for anomalies published before 1995.

There are improvements in mean returns and Sharpe ratios for both the equal-weighted and value-weighted portfolios in the US with addition of the new anomalies. The new anomalies

[29] For example heavily cited size and book-to-value factor in Fama and French (1992) were introduced before 1990.

[30] Adding another set of anomalies published before 2010 and focusing on 2010-2016 out-of-sample period leads to identical findings. The corresponding results are available in table IA4 in the online appendix.

Table XIII
Are the More Recent Anomalies Improving Profitability of the Mispricing Strategy?

The table shows returns of the mispricing strategy described in Table IV that is estimated on the individual stocks from the US. Anomalies in the estimation are restricted to those that were published before 1995, 2000, or 2005. The returns are reported in percentage points per month over the 2005-2016 period.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Weighted Least Squares										
Published by 1995										
Mean	0.044	0.082	0.361	0.294	0.169	0.268	0.297	0.914	0.284	0.256
Sharpe Ratio	0.034	0.068	0.266	0.194	0.171	0.210	0.226	0.615	0.155	0.234
Published by 2000										
Mean	0.055	0.278	0.341	-0.106	0.164	0.119	0.306	0.595	-0.238	0.232
Sharpe Ratio	0.047	0.237	0.234	-0.072	0.177	0.093	0.241	0.363	-0.140	0.234
Information Ratio	0.016	0.275	-0.025	-0.374	-0.010	-0.180	0.011	-0.326	-0.394	-0.038
Published by 2005										
Mean	0.612	0.078	0.854	0.270	0.506	0.685	0.105	1.087	-0.004	0.556
Sharpe Ratio	0.564	0.074	0.646	0.189	0.624	0.644	0.087	0.787	-0.002	0.684
Information Ratio	0.856	-0.265	0.702	0.291	0.724	0.545	-0.173	0.482	0.134	0.444
Gradient Boosting Regression Trees										
Published by 1995										
Mean	0.364	0.276	0.868	0.816	0.531	0.411	0.192	0.917	0.197	0.289
Sharpe Ratio	0.369	0.216	0.772	0.521	0.692	0.377	0.133	0.766	0.114	0.322
Published by 2000										
Mean	0.435	0.647	0.979	1.058	0.722	0.250	0.476	0.951	0.943	0.430
Sharpe Ratio	0.408	0.483	0.873	0.727	0.901	0.234	0.313	0.720	0.602	0.461
Information Ratio	0.091	0.453	0.122	0.226	0.430	-0.181	0.267	0.026	0.477	0.227
Published by 2005										
Mean	0.824	0.602	1.212	1.043	0.904	0.948	0.381	1.138	0.414	0.777
Sharpe Ratio	0.842	0.537	1.121	0.864	1.309	1.012	0.332	0.980	0.314	1.135
Information Ratio	0.585	-0.054	0.276	-0.012	0.403	0.819	-0.090	0.169	-0.349	0.515

therefore have significant incremental value for out-of-sample forecasts. This benefit is smaller in Japan and Europe. The results are similar for both least squares and gradient boosting regression trees methods but the returns from least squares are much more volatile. One explanation for the larger incremental value of the new anomalies in the US with respect to Europe and Japan is that there are more low-cost exchange traded funds in the US that arbitrage away the well-known strategies. It is therefore necessary to find new strategies to get the same predictability of stock returns over time.

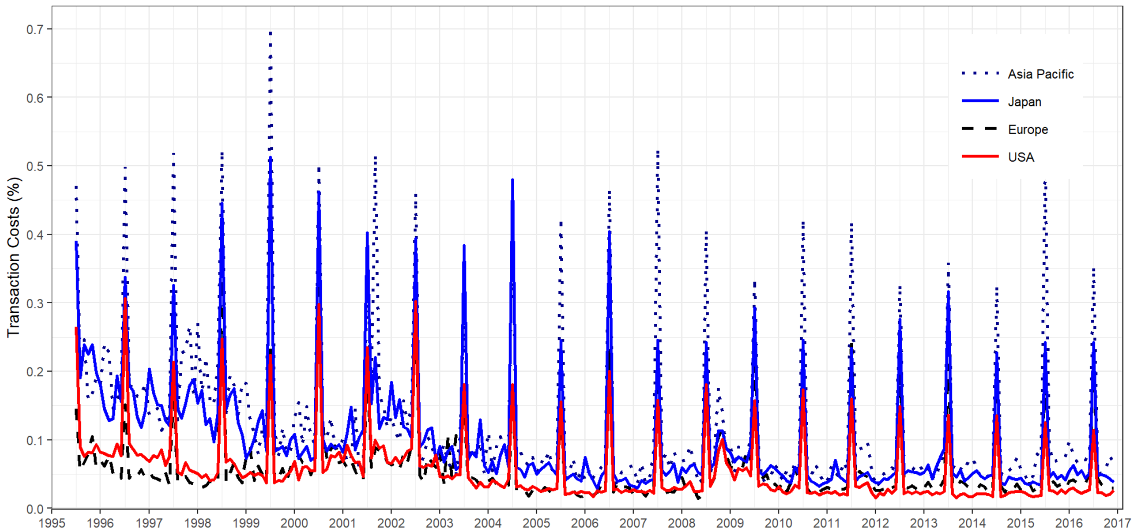
To conclude, the marginal value of the new anomalies remains positive over time. It is therefore valuable to follow recent academic research as it can increase returns to investors. The positive value of new anomalies is in line with the purpouse of academic publishing process where new findings are put under scrutiny and the authors have to prove that their findings provide incremental value with respect to the existing body of knowledge. The academic review process therefore fulfills its purpouse.

V. Transaction Costs

This section studies the out-of-sample performance of the strategies after the transaction costs. It is possible that the profits on the strategies are only virtual and transaction costs are larger than the returns. It is therefore important to examine the costs related to the strategies.

A. *Transaction Costs on the Strategies*

Panel A: Portfolio-mixing Strategy.



Panel B: Gradient Boosting Regression Trees Mispricing Strategy.

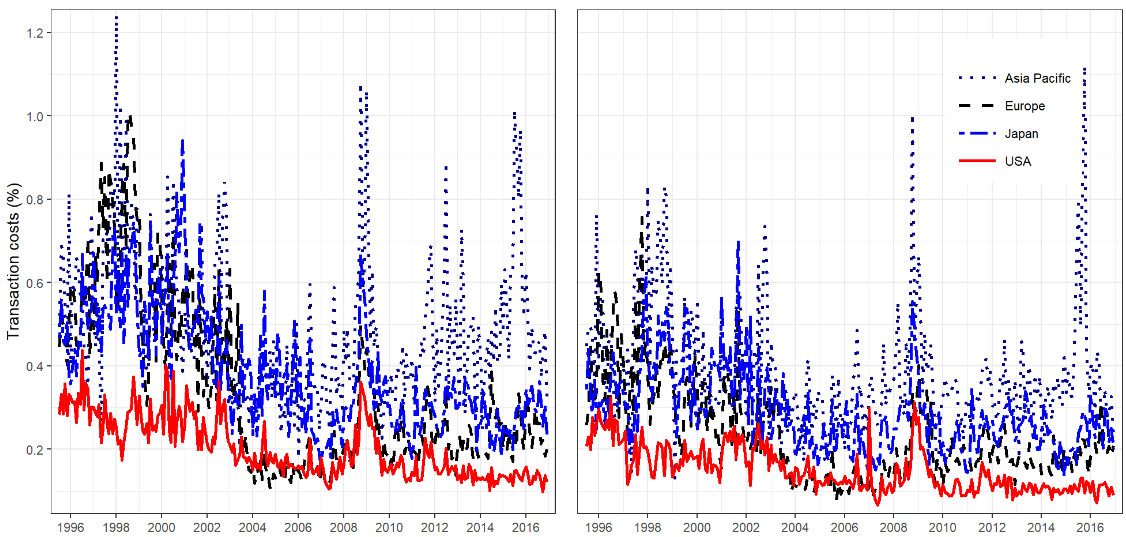


Figure 9. Monthly Transaction Costs. Panel A shows transaction costs for the portfolio mixing strategy that equally invests in all the significant anomalies as described in Table III. Panel B shows transaction costs for the mispricing strategy described in Table IV that is estimated on individual stock returns from the US. The transaction costs are estimated with VoV(% Spread) proxy of Fong et al. (2017).

Panel A in Figure 9 describes transaction costs on the portfolio-mixing strategy introduced in

Section II.A. The transaction costs are measured by VoV(% Spread) proxy introduced in Fong et al. (2017). It is evident that the trading costs are similar across the regions for the liquid sample of stocks. The highest transaction costs tend to be in Asia Pacific region. The peaks in the figure appear every July because of the annual rebalancing of the fundamental strategies. The graph also documents that there are periods with significant spillover of illiquidity. Two such major episodes are Global Financial Crisis of 2008 and Dot-com bubble of early 2000s. The transaction costs have decreased significantly over time with the increase in market share of electronic trading in 2000s.

The transaction costs on the mispricing strategy are covered next. Panel B in Figure 9 maps transaction costs on the gradient boosting regression trees strategy estimated in the US. It is apparent that the transaction costs are larger than in case of portfolio-mixing strategy. The costs are larger because a large portion of the individual anomalies are fundamental anomalies that are rebalanced annually, whereas, the mispricing strategy is rebalanced monthly. The transaction costs have decreased significantly over time and there are again several historical episodes where they were heavily elevated, one being the Global Financial Crisis of 2008. The costs are smaller on value-weighted portfolios relative to equal-weighting which is expected because the value-weighting puts larger weight on more liquid stocks.

Table XIV
Transaction Costs on the Mispricing Strategy

The table shows transaction costs and turnover on the gradient boosting regression trees mispricing strategy described in Table IV that is estimated on the individual stocks from the US. The transaction costs are estimated either with VoV(% Spread) proxy of Fong et al. (2017), average daily closing quoted spread, or Gibbs proxy of Hasbrouck (2009). The transaction costs and turnover are in percentage points per month.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
VoV	0.203	0.343	0.388	0.493	0.357	0.151	0.232	0.288	0.386	0.264
Gibbs	0.819	0.651	0.816	0.784	0.767	0.712	0.567	0.733	0.714	0.681
Quoted Spread	0.111	0.511	0.482	0.791	0.473	0.101	0.384	0.418	0.645	0.387
Turnover	120.0	119.2	118.9	123.6	120.4	130.5	127.1	127.7	139.4	131.2

Table XIV presents average transaction costs on the gradient boosting regression trees mispricing strategy. The transaction costs are estimated with three liquidity proxies introduced in section I.E. All the proxies provide very similar estimates of the transaction costs outside the US. Estimates from Gibbs proxy are significantly higher in the US than for the two other proxies. Gibbs proxy is, however, also the most noisy proxy since it is constructed at an annual frequency. It is furthermore not very suitable to measure transaction costs for the most liquid stocks due to its construction.

Table XIV also shows turnover of the mispricing strategy. The turnover is defined as

$$Turnover_t = \sum_i abs(w_{i,t} - w_{i,t-1}r_{i,t-1})/2 \tag{5}$$

where $w_{i,t}$ is weight of stock i in the investment portfolio at the start of period $t-1$ and $r_{i,t-1}$ is stock return over period $t-1$ to t. Sum of all absolute weights $w_{i,t}$ is equal to 2 since the portfolio is long-short. The turnover is close to 125% monthly in all the regions, which means that over 60% of all the held stocks have to be sold and new bought for both the short and long leg of the strategy. The turnover can be easily reduced by staggered portfolio rebalancing but it is not a source of serious worries here due to the small average transaction costs on the liquid universe of stocks.

The sample of stocks has been selected to be liquid ex ante. Only about 500 most liquid US stocks fulfill this criterion. These stocks should be with virtually no fixed transaction costs. The depicted costs therefore correspond to unfavorable trade executions through aggressive marketable orders. Sophisticated trade execution systems using limit orders are able to execute the strategies without any transaction costs.

B. Performance of the Strategies after Transaction Costs

B.1. Portfolio-mixing Strategy

Panel A in Table XV presents returns on the portfolio-mixing strategy introduced in Table III adjusted for the trading costs. The set of selected significant strategies is different from Table III as the strategies are selected on after cost basis here. The selection after adjusting for transaction costs leads to a more profitable meta-strategy as the anomalies with the largest profitability are also often those with the largest transaction costs.

Returns on the strategy remain positive outside Japan but they are generally smaller than without the transaction costs. The Sharpe ratios are also smaller. The global portfolio-mixing strategy, however, remains significantly profitable with Sharpe ratio close to 0.5 for value-weighted returns.

B.2. Mispricing Strategy

Panel B in Table XV presents performance of the mispricing strategy after transaction costs. The mean returns on the strategy remain significantly positive at 5% level. The net mean annualized returns in the US are above 10% for the machine learning strategies. Sharpe ratios remain high, especially for the global strategy using neural networks where they are larger than one.

The mean returns after transaction costs for weighted least square method are again smaller

Table XV
Performance of the Strategies after Transaction Costs

Panel A shows returns minus transaction costs of the portfolio-mixing strategy described in Table III. Panel B shows returns after transaction costs of the mispricing strategy described in Table IV that is estimated on the individual stocks from the US. The transaction costs are estimated with VoV(% Spread) proxy of Fong et al. (2017). The returns are reported in percentage points per month.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Panel A: Portfolio-mixing Strategy										
Mean	0.151	0.225	-0.119	0.612	0.217	0.109	0.066	0.185	0.589	0.237
Sharpe Ratio	0.186	0.340	-0.165	0.609	0.399	0.184	0.104	0.202	0.630	0.488
Skewness	0.272	0.024	-2.562	-1.255	-0.630	-0.044	0.202	0.260	1.264	0.643
Kurtosis	8.958	7.277	20.36	13.26	6.979	5.907	9.065	23.91	13.59	16.12
Max Drawdown	-29.47	-23.33	-42.78	-28.47	-15.07	-22.87	-34.18	-48.59	-20.75	-11.80
Panel B: Mispricing Strategy										
Weighted Least Squares										
Mean	0.583	0.336	0.514	0.269	0.439	0.583	0.336	0.514	0.269	0.439
Sharpe Ratio	0.348	0.267	0.390	0.170	0.412	0.249	0.294	0.217	0.124	0.313
Max Drawdown	-66.74	-44.12	-49.35	-58.25	-49.63	-71.40	-39.33	-59.36	-58.92	-55.91
Penalized Weighted Least Squares										
Mean	0.537	0.379	0.471	0.344	0.426	0.480	0.547	0.284	0.358	0.403
Sharpe Ratio	0.315	0.290	0.353	0.211	0.385	0.284	0.381	0.177	0.173	0.335
Max Drawdown	-67.37	-43.85	-48.82	-54.35	-51.57	-69.77	-38.97	-58.36	-65.62	-54.72
Gradient Boosting Regression Trees										
Mean	0.962	0.527	0.785	1.157	0.806	1.240	0.359	0.723	1.029	0.769
Sharpe Ratio	0.594	0.390	0.513	0.704	0.793	0.741	0.250	0.376	0.581	0.648
Max Drawdown	-39.92	-49.42	-36.00	-41.48	-29.34	-44.94	-45.90	-48.05	-41.42	-37.52
Random Forest										
Mean	0.844	0.681	0.714	0.414	0.703	0.825	0.089	0.751	0.770	0.525
Sharpe Ratio	0.565	0.513	0.504	0.228	0.706	0.584	0.058	0.431	0.398	0.477
Max Drawdown	-34.62	-49.52	-41.89	-63.52	-29.97	-31.22	-61.80	-44.32	-46.17	-32.73
Neural Networks										
Mean	1.222	0.785	0.934	1.296	1.016	1.282	0.610	0.834	0.829	0.851
Sharpe Ratio	0.782	0.630	0.815	0.804	1.195	0.818	0.479	0.526	0.492	0.812
Max Drawdown	-46.21	-35.29	-25.93	-41.99	-20.70	-49.19	-39.00	-38.28	-58.37	-35.83

than for the more advanced machine learning methods. The difference is even larger on risk adjusted basis. This difference in performance documents that the choice of appropriate forecasting method is very important for success of investing into the anomalies.

To conclude, the strategies remain profitable even after accounting for the transaction costs. The profitability of the strategies is therefore not illusory and can be capitalized by the investors.

VI. Conclusion

This study has examined profitability of the quantitative strategies based on published anomalies around the globe. It has been shown that investing into individual anomalies is profitable after accounting for transaction costs even on liquid universe of stocks. The performance of the strategy combining individual portfolios on anomalies can be improved by creating a single mispricing signal instead. Machine learning approach for construction of the mispricing signal was advocated and its benefits documented.

The machine learning methods lead to higher (risk adjusted) returns relative to standard methods applied in the academic finance literature. The quantitative strategy using machine learning is highly profitable even on liquid universe of stocks. Value of the more recent anomalies was then studied. The recently published anomalies improve average returns on the investment strategy even after accounting for the previously published anomalies. The recent anomaly studies are therefore successful in finding new sources of priced risk and investors' behavioural biases.

The role of international evidence on precision of predictions of future stock returns was studied. Out-of-sample performance in the US is not improved with international evidence in the training sample for the mispricing strategy. Most of the predictability of expected stock returns in all the global regions under study can be captured solely with the US training sample.

Appendix A. Adjustments of Returns in Datastream

A series of adjustments is applied on the raw returns to improve their quality. Return index (RI) is required to be larger than 0.001 on the first day of the month for precision reasons. RI is set to missing if daily return is larger than 500% or if price on the first day of the month is larger than $1 million. Any monthly return larger than 2000% is also set to missing. Datastream provides stale prices when there is no trade during the day or when the stock is no longer traded so that the price of the last trade is repeated until new information arrives. The latest observations of price with no trading are therefore deleted. Daily returns are fixed following Tobek and Hronec (2018) when there are stale price quotes around corporate events. Monthly returns larger than 300% that revert back over the next month are set to missing following Ince and Porter (2006).[31] 0.01% of returns is winsorized in each region and year before 2000 to limit role of outliers in returns.

Appendix B. List of the Anomalies

Table XVI
List of Anomalies

Fundamental	
Accruals	
Accruals	Sloan (1996)
Change in Common Equity	Richardson, Sloan, Soliman, and Tuna (2006)
Change in Current Operating Assets	Richardson et al. (2006)
Change in Current Operating Liabilities	Richardson et al. (2006)

[31]Specifically, returns in two consecutive months are set as missing if the return in the first month is larger than 300% and the overall return over the two months is lower than 50%.

Change in Financial Liabilities	Richardson et al. (2006)
Change in Long-Term Investments	Richardson et al. (2006)
Change in Net Financial Assets	Richardson et al. (2006)
Change in Net Non-Cash Working Capital	Richardson et al. (2006)
Change in Net Non-Current Operating Assets	Richardson et al. (2006)
Change in Non-Current Operating Assets	Richardson et al. (2006)
Change in Non-Current Operating Liabilities	Richardson et al. (2006)
Change in Short-Term Investments	Richardson et al. (2006)
Discretionary Accruals	Dechow, Sloan, and Sweeney (1995)
Growth in Inventory	Thomas and Zhang (2002)
Inventory Change	Thomas and Zhang (2002)
Inventory Growth	Belo and Lin (2011)
M/B and Accruals	Bartov and Kim (2004)
Net Working Capital Changes	Soliman (2008)
Percent Operating Accrual	Hafzalla, Lundholm, and Matthew Van Winkle (2011)
Percent Total Accrual	Hafzalla et al. (2011)
Total Accruals	Richardson et al. (2006)
Intangibles	
$\triangle$ Gross Margin - $\triangle$ Sales	Abarbanell and Bushee (1998)
$\triangle$ Sales - $\triangle$ Accounts Receivable	Abarbanell and Bushee (1998)
$\triangle$ Sales - $\triangle$ Inventory	Abarbanell and Bushee (1998)
$\triangle$ Sales - $\triangle$ SG and A	Abarbanell and Bushee (1998)
Asset Liquidity	Ortiz-Molina and Phillips (2014)
Asset Liquidity II	Ortiz-Molina and Phillips (2014)
Cash-to-assets	Palazzo (2012)
Earnings Conservatism	Francis, LaFond, Olsson, and Schipper (2004)
Earnings Persistence	Francis et al. (2004)
Earnings Predictability	Francis et al. (2004)
Earnings Smoothness	Francis et al. (2004)
Earnings Timeliness	Francis et al. (2004)
Herfindahl Index	Hou and Robinson (2006)
Hiring rate	Belo, Lin, and Bazdresch (2014)
Industry Concentration Assets	Hou and Robinson (2006)
Industry Concentration Book Equity	Hou and Robinson (2006)
Industry-adjusted Organizational Capital-to-Assets	Eisfeldt and Papanikolaou (2013)
Industry-adjusted Real Estate Ratio	Tuzel (2010)
Org. Capital	Eisfeldt and Papanikolaou (2013)
RD / Market Equity	Chan et al. (2001)
RD Capital-to-assets	Li (2011)
RD Expenses-to-sales	Chan et al. (2001)
Tangibility	Hahn and Lee (2009)
Unexpected RD Increases	Eberhart, Maxwell, and Siddique (2004)
Whited-Wu Index	Whited and Wu (2006)
Investment	

△ CAPEX - △ Industry CAPEX	Abarbanell and Bushee (1998)
Asset Growth	Cooper, Gulen, and Schill (2008)
Change Net Operating Assets	Hirshleifer, Hou, Teoh, and Zhang (2004)
Changes in PPE and Inventory-to-Assets	Lyandres, Sun, and Zhang (2007)
Composite Debt Issuance	Lyandres et al. (2007)
Composite Equity Issuance (5-Year)	Daniel and Titman (2006)
Debt Issuance	Spiess and Affleck-Graves (1995)
Growth in LTNOA	Fairfield, Whisenant, and Yohn (2003)
Investment	Titman, Wei, and Xie (2004)
Net Debt Finance	Bradshaw, Richardson, and Sloan (2006)
Net Equity Finance	Bradshaw et al. (2006)
Net Operating Assets	Hirshleifer et al. (2004)
Noncurrent Operating Assets Changes	Soliman (2008)
Share Repurchases	Ikenberry, Lakonishok, and Vermaelen (1995)
Total XFIN	Bradshaw et al. (2006)
Profitability	
Asset Turnover	Soliman (2008)
Capital Turnover	Haugen and Baker (1996)
Cash-based Operating Profitability	Ball, Gerakos, Linnainmaa, and Nikolaev (2016)
Change in Asset Turnover	Soliman (2008)
Change in Profit Margin	Soliman (2008)
Earnings / Price	Basu (1977)
Earnings Consistency	Alwathainani (2009)
F-Score	Piotroski (2000)
Gross Profitability	Novy-Marx (2013)
Labor Force Efficiency	Abarbanell and Bushee (1998)
Leverage	Bhandari (1988)
O-Score (More Financial Distress)	Dichev (1998)
Operating Profits to Assets	Ball et al. (2016)
Operating Profits to Equity	Fama and French (2015)
Profit Margin	Soliman (2008)
Return on Net Operating Assets	Soliman (2008)
Return-on-Equity	Haugen and Baker (1996)
Z-Score (Less Financial Distress)	Dichev (1998)
Value	
Assets-to-Market	Fama and French (1992)
Book Equity / Market Equity	Fama and French (1992)
Cash Flow / Market Equity	Lakonishok, Shleifer, and Vishny (1994)
Duration of Equity	Dechow, Sloan, and Soliman (2004)
Enterprise Component of Book/Price	Penman, Richardson, and Tuna (2007)
Enterprise Multiple	Loughran and Wellman (2011)
Intangible Return	Daniel and Titman (2006)
Leverage Component of Book/Price	Penman et al. (2007)
Net Payout Yield	Boudoukh, Michaely, Richardson, and Roberts (2007)

Operating Leverage	Novy-Marx (2010)
Payout Yield	Boudoukh et al. (2007)
Sales Growth	Lakonishok et al. (1994)
Sales/Price	Barbee Jr, Mukherji, and Raines (1996)
Sustainable Growth	Lockwood and Prombutr (2010)
Market Friction	
11-Month Residual Momentum	Blitz, Huij, and Martens (2011)
52-Week High	George and Hwang (2004)
Amihud's Measure (Illiquidity)	Amihud (2002)
Beta	Fama and MacBeth (1973)
Betting against Beta	Frazzini and Pedersen (2014)
Bid-Ask Spread	Amihud and Mendelson (1986)
Cash Flow Variance	Haugen and Baker (1996)
Coefficient of Variation of Share Turnover	Chordia, Subrahmanyam, and Anshuman (2001)
Coskewness	Harvey and Siddique (2000)
Downside Beta	Ang, Chen, and Xing (2006a)
Earnings Forecast-to-Price	Elgers, Lo, and Pfeiffer Jr (2001)
Firm Age	Barry and Brown (1984)
Firm Age-Momentum	Zhang (2006)
Idiosyncratic Risk	Ang, Hodrick, Xing, and Zhang (2006b)
Industry Momentum	Moskowitz and Grinblatt (1999)
Lagged Momentum	Novy-Marx (2012)
Liquidity Beta 1	Acharya and Pedersen (2005)
Liquidity Beta 2	Acharya and Pedersen (2005)
Liquidity Beta 3	Acharya and Pedersen (2005)
Liquidity Beta 4	Acharya and Pedersen (2005)
Liquidity Beta 5	Acharya and Pedersen (2005)
Liquidity Shocks	Bali, Peng, Shen, and Tang (2013)
Long-Term Reversal	Bondt and Thaler (1985)
Max	Bali, Cakici, and Whitelaw (2011)
Momentum	Jegadeesh and Titman (1993)
Momentum and LT Reversal	Kot and Chan (2006)
Momentum-Reversal	Jegadeesh and Titman (1993)
Momentum-Volume	Lee and Swaminathan (2000)
Price	Blume and Husic (1973)
Seasonality	Heston and Sadka (2008)
Seasonality 1 A	Heston and Sadka (2008)
Seasonality 1 N	Heston and Sadka (2008)
Seasonality 11-15 A	Heston and Sadka (2008)
Seasonality 11-15 N	Heston and Sadka (2008)
Seasonality 16-20 A	Heston and Sadka (2008)
Seasonality 16-20 N	Heston and Sadka (2008)
Seasonality 2-5 A	Heston and Sadka (2008)
Seasonality 2-5 N	Heston and Sadka (2008)

Seasonality 6-10 A	Heston and Sadka (2008)
Seasonality 6-10 N	Heston and Sadka (2008)
Share Issuance (1-Year)	Pontiff and Woodgate (2008)
Share Turnover	Datar, Naik, and Radcliffe (1998)
Short-Term Reversal	Jegadeesh (1990)
Size	Banz (1981)
Tail Risk	Kelly and Jiang (2014)
Total Volatility	Ang et al. (2006b)
Volume / Market Value of Equity	Haugen and Baker (1996)
Volume Trend	Haugen and Baker (1996)
Volume Variance	Chordia et al. (2001)
I/B/E/S	
Analyst Value	Frankel and Lee (1998)
Analysts Coverage	Elgers et al. (2001)
Change in Forecast + Accrual	Barth and Hutton (2004)
Change in Recommendation	Jegadeesh, Kim, Krische, and Lee (2004)
Changes in Analyst Earnings Forecasts	Hawkins, Chamberlin, and Daniel (1984)
Disparity between LT and ST Earnings Growth Forecasts	Da and Warachka (2011)
Dispersion in Analyst LT Growth Forecasts	Anderson, Ghysels, and Juergens (2005)
Down Forecast	Barber, Lehavy, McNichols, and Trueman (2001)
Forecast Dispersion	Diether, Malloy, and Scherbina (2002)
Long-Term Growth Forecasts	La Porta (1996)
Up Forecast	Barber et al. (2001)

Appendix C. Definition of Liquidity Proxies

Appendix A. VoV(% Spread) Proxy

The fixed transaction costs are approximated with VoV(% Spread) proxy introduced in Fong et al. (2017). It is defined as

$$8\,\frac{\sigma^{2/3}}{avg\ vol^{1/3}} \tag{C1}$$

where σ is standard deviation of daily returns and *avg vol* is average daily trading volume in USD within a given month. The trading volume is in USD and deflated to 2000 prices. The proxy roughly measures fixed component of trading costs and excludes price impact. Including the price impact would further increase the transaction costs. Fong et al. (2017) show that the price impact component is very hard to measure. It is volatile over regions, and therefore, very dependent on execution strategy of individual asset managers. The focus is therefore solely on the fixed component of transaction costs (effective spread).

Kyle and Obizhaeva (2016) estimated a relationship between transaction costs and size of large institutional portfolio transfers depending on average daily trading volume and volatility of the stocks. The analysis was conducted on a proprietary dataset covering the 2002-2005 period. VoV(% Spread) roughly corresponds to the fixed component of their estimated transaction cost function.

Fong et al. (2017) benchmarked the proxy to other existing proxies and found that it can be outperformed only by closing quoted spread. The quoted spread is, however, not available for all the regions over the whole sample period.

Appendix B. Closing Quoted Spread

Closing quoted spread for a given month is defined as

$$QS = \frac{1}{T}\sum_{t=1}^{T}\frac{2(ask-bid)}{ask+bid} \tag{C2}$$

where ask and bid are observed at the end of trading day on each stock exchange and T is number of days in the given month. Observations with missing or negative daily value of QS are excluded from the average. CRSP lists the best quote of bid and ask for NASDAQ stocks and the last representative quotes before the market close for NYSE and Amex stocks. Precise definition of QS can therefore vary over the exchanges.

Chung and Zhang (2014) first benchmarked the QS by comparing it to high frequency effective spread estimates from Trade and Quote (TAQ) database. They showed that QS has about 95% average cross sectional correlation with TAQ effective spread over the 1998 to 2009 period. Fong et al. (2017) document that it is also the best spread proxy in international setting. One problem with QS is that it is often missing in earlier periods and therefore has to be backfilled with other proxies.

Appendix C. Gibbs Proxy

Roll (1984) introduced one of the first spread proxies in the academic literature. He assumed that the true price of stock follows a random walk with bid-ask jumps. That is,

$$P_t^A = P_{t-1}^A + u_t, \quad P_t^O = P_t^A + sq_t \tag{C3}$$

$$\triangle P_t^o = s \triangle q_t + u_t,\, u_t \sim N(0, \sigma_u^2) \tag{C4}$$

where P_t^o is observed log price, P_t^A is price of the underlying Brownian motion, and s is a half spread. Indicator q_t is equal to one if the last trade in the day is buy, minus one if it is sell, and zero if no prices are available during the day. Serial correlation of the price changes $\triangle P_t^o$ should be negative and related to the spread through the following relationship

$$S_{roll} = 2\sqrt{-cov(\triangle P_t^o, \triangle P_{t+1}^o)}. \tag{C5}$$

This can be contributed to the fact that

$$cov(\triangle P_t^o, \triangle P_{t+1}^o) = cov(s(q_t - q_{t-1}) + u_t, s(q_{t+1} - q_t) + u_{t+1}) = \mathbb{E}[-s^2 q_t^2] = -s^2. \tag{C6}$$

The covariance can be positive in practice. In which case the estimate of spread is set equal to zero.

Hasbrouck (2009) proposed to extend the Roll model by estimating it with Gibbs sampler. The idea is to estimate the equation (C4) augmented with another dependent variable (market return) via Bayesian regression. The variables q_t are generated from the data by Gibbs sampler.[32]

The proxy is estimated at annual frequency for each stock and calendar year. Lower frequency than annual leads to severe deterioration of the proxy's performance.

[32] Note that there is an error in the original paper in *Journal of Finance*. The correct posterior distribution for σ_u^2 is $IG(\alpha_{prior} + \frac{n}{2}, \beta_{prior} + \frac{\sum u_t^2}{2})$.

References

Abarbanell, Jeffery S, and Brian J Bushee, 1998, Abnormal returns to a fundamental analysis strategy, *Accounting Review* 19–45.

Acharya, Viral V, and Lasse Heje Pedersen, 2005, Asset pricing with liquidity risk, *Journal of financial Economics* 77, 375–410.

Alwathainani, Abdulaziz M, 2009, Consistency of firms' past financial performance measures and future returns, *The British Accounting Review* 41, 184–196.

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of financial markets* 5, 31–56.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of financial Economics* 17, 223–249.

Anderson, Evan W, Eric Ghysels, and Jennifer L Juergens, 2005, Do heterogeneous beliefs matter for asset pricing?, *The Review of Financial Studies* 18, 875–924.

Andrikopoulos, Panagiotis, James Clunie, and Antonios Siganos, 2013, Short-selling constraints and 'quantitative'investment strategies, *The European Journal of Finance* 19, 19–35.

Ang, Andrew, Joseph Chen, and Yuhang Xing, 2006a, Downside risk, *The Review of Financial Studies* 19, 1191–1239.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006b, The cross-section of volatility and expected returns, *The Journal of Finance* 61, 259–299.

Asness, Clifford, and Andrea Frazzini, 2013, The devil in hml's details, *The Journal of Portfolio Management* 39, 49–68.

Asparouhova, Elena, Hendrik Bessembinder, and Ivalina Kalcheva, 2010, Liquidity biases in asset pricing tests, *Journal of Financial Economics* 96, 215–237.

Bali, Turan G, Nusret Cakici, and Robert F Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99, 427–446.

Bali, Turan G, Lin Peng, Yannan Shen, and Yi Tang, 2013, Liquidity shocks and stock market reactions, *The Review of Financial Studies* 27, 1434–1485.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, and Valeri Nikolaev, 2016, Accruals, cash flows, and operating profitability in the cross section of stock returns, *Journal of Financial Economics* 121, 28–45.

Banz, Rolf W, 1981, The relationship between return and market value of common stocks, *Journal of financial economics* 9, 3–18.

Barbee Jr, William C, Sandip Mukherji, and Gary A Raines, 1996, Do sales–price and debt–equity explain stock returns better than book–market and firm size?, *Financial Analysts Journal* 52, 56–60.

Barber, Brad, Reuven Lehavy, Maureen McNichols, and Brett Trueman, 2001, Can investors profit from the prophets? security analyst recommendations and stock returns, *The Journal of Finance* 56, 531–563.

Barber, Brad M, Emmanuel T De George, Reuven Lehavy, and Brett Trueman, 2013, The earnings announcement premium around the globe, *Journal of Financial Economics* 108, 118–138.

Barry, Christopher B, and Stephen J Brown, 1984, Differential information and the small firm effect, *Journal of Financial Economics* 13, 283–294.

Barth, Mary E, and Amy P Hutton, 2004, Analyst earnings forecast revisions and the pricing of accruals, *Review of accounting studies* 9, 59–96.

Bartov, Eli, and Myungsun Kim, 2004, Risk, mispricing, and value investing, *Review of Quantitative Finance and Accounting* 23, 353–376.

Bartram, Söhnke M, and Mark Grinblatt, 2018, Global market inefficiencies .

Basu, Sanjoy, 1977, Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis, *The journal of Finance* 32, 663–682.

Belo, Frederico, and Xiaoji Lin, 2011, The inventory growth spread, *The Review of Financial Studies* 25, 278–313.

Belo, Frederico, Xiaoji Lin, and Santiago Bazdresch, 2014, Labor hiring, investment, and stock return predictability in the cross section, *Journal of Political Economy* 122, 129–177.

Bhandari, Laxmi Chand, 1988, Debt/equity ratio and expected common stock returns: Empirical evidence, *The journal of finance* 43, 507–528.

Blitz, David, Joop Huij, and Martin Martens, 2011, Residual momentum, *Journal of Empirical Finance* 18, 506–521.

Blume, Marshall E, and Frank Husic, 1973, Price, beta, and exchange listing, *The Journal of Finance* 28, 283–299.

Bondt, Werner FM, and Richard Thaler, 1985, Does the stock market overreact?, *The Journal of finance* 40, 793–805.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *The Journal of Finance* 62, 877–915.

Bradshaw, Mark T, Scott A Richardson, and Richard G Sloan, 2006, The relation between corporate financing activities, analysts' forecasts and stock returns, *Journal of Accounting and Economics* 42, 53–85.

Chan, Louis KC, Josef Lakonishok, and Theodore Sougiannis, 2001, The stock market valuation of research and development expenditures, *The Journal of Finance* 56, 2431–2456.

Chen, Luyang, Markus Pelger, and Jason Zhu, 2019, Deep learning in asset pricing, *Available at SSRN 3350138* .

Chen, Tianqi, and Tong He, 2017, xgboost: extreme gradient boosting .

Chordia, Tarun, Avanidhar Subrahmanyam, and V Ravi Anshuman, 2001, Trading activity and expected stock returns, *Journal of Financial Economics* 59, 3–32.

Chui, Andy CW, Sheridan Titman, and KC John Wei, 2010, Individualism and momentum around the world, *The Journal of Finance* 65, 361–392.

Chung, Kee H, and Hao Zhang, 2014, A simple approximation of intraday spreads using daily data, *Journal of Financial Markets* 17, 94–120.

Cooper, Michael J, Huseyin Gulen, and Michael J Schill, 2008, Asset growth and the cross-section of stock returns, *The Journal of Finance* 63, 1609–1651.

Da, Zhi, and Mitch Warachka, 2011, The disparity between long-term and short-term forecasted earnings growth, *Journal of Financial Economics* 100, 424–442.

Daniel, Kent, and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *The Journal of Finance* 61, 1605–1643.

Datar, Vinay T, Narayan Y Naik, and Robert Radcliffe, 1998, Liquidity and stock returns: An alternative test, *Journal of Financial Markets* 1, 203–219.

Dechow, Patricia M, Richard G Sloan, and Mark T Soliman, 2004, Implied equity duration: A new measure of equity risk, *Review of Accounting Studies* 9, 197–228.

Dechow, Patricia M, Richard G Sloan, and Amy P Sweeney, 1995, Detecting earnings management, *Accounting review* 193–225.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2007, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *The review of Financial studies* 22, 1915–1953.

Dichev, Ilia D, 1998, Is the risk of bankruptcy a systematic risk?, *the Journal of Finance* 53, 1131–1147.

Diether, Karl B, Christopher J Malloy, and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, *The Journal of Finance* 57, 2113–2141.

Driscoll, John C, and Aart C Kraay, 1998, Consistent covariance matrix estimation with spatially dependent panel data, *Review of economics and statistics* 80, 549–560.

Eberhart, Allan C, William F Maxwell, and Akhtar R Siddique, 2004, An examination of long-term abnormal stock returns and operating performance following r&d increases, *The Journal of Finance* 59, 623–650.

Eisfeldt, Andrea L, and Dimitris Papanikolaou, 2013, Organization capital and the cross-section of expected returns, *The Journal of Finance* 68, 1365–1406.

Elgers, Pieter T, May H Lo, and Ray J Pfeiffer Jr, 2001, Delayed security price adjustments to financial analysts' forecasts of annual earnings, *The Accounting Review* 76, 613–632.

Fairfield, Patricia M, J Scott Whisenant, and Teri Lombardi Yohn, 2003, Accrued earnings and growth: Implications for future profitability and market mispricing, *The accounting review* 78, 353–371.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *the Journal of Finance* 47, 427–465.

Fama, Eugene F, and Kenneth R French, 2012, Size, value, and momentum in international stock returns, *Journal of financial economics* 105, 457–472.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1 – 22.

Fama, Eugene F, and Kenneth R French, 2017, International tests of a five-factor asset pricing model, *Journal of Financial Economics* 123, 441–463.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Fong, Kingsley, Craig Holden, and Ondrej Tobek, 2017, Are volatility over volume liquidity proxies useful for global or us research? .

Francis, Jennifer, Ryan LaFond, Per M Olsson, and Katherine Schipper, 2004, Costs of equity and earnings attributes, *The accounting review* 79, 967–1010.

Frankel, Richard, and Charles MC Lee, 1998, Accounting valuation, market expectation, and cross-sectional stock returns, *Journal of Accounting and economics* 25, 283–319.

Frazzini, Andrea, Ronen Israel, and Tobias J Moskowitz, 2012, Trading costs of asset pricing anomalies, *Fama-Miller working paper* 14–05.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, 2001, *The elements of statistical learning*, volume 1 (Springer series in statistics New York, NY, USA:).

Friedman, Jerome H, 2001, Greedy function approximation: a gradient boosting machine, *Annals of statistics* 1189–1232.

George, Thomas J, and Chuan-Yang Hwang, 2004, The 52-week high and momentum investing, *The Journal of Finance* 59, 2145–2176.

Green, Jeremiah, John RM Hand, and X Frank Zhang, 2017, The characteristics that provide independent information about average us monthly stock returns, *The Review of Financial Studies* hhx019.

Griffin, John M, 2002, Are the fama and french factors global or country specific?, *The Review of Financial Studies* 15, 783–803.

Griffin, John M, Patrick J Kelly, and Federico Nardari, 2010, Do market efficiency measures yield correct inferences? a comparison of developed and emerging markets, *The Review of Financial Studies* 23, 3225–3277.

Gu, Shihao, Bryan T Kelly, and Dacheng Xiu, 2018, Empirical asset pricing via machine learning .

Hafzalla, Nader, Russell Lundholm, and E Matthew Van Winkle, 2011, Percent accruals, *The Accounting Review* 86, 209–236.

Hahn, Jaehoon, and Hangyong Lee, 2009, Financial constraints, debt capacity, and the cross-section of stock returns, *The Journal of Finance* 64, 891–921.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2016, . . . and the cross-section of expected returns, *The Review of Financial Studies* 29, 5–68.

Harvey, Campbell R, and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *The Journal of Finance* 55, 1263–1295.

Hasbrouck, Joel, 2009, Trading costs and returns for us equities: Estimating effective costs from daily data, *The Journal of Finance* 64, 1445–1477.

Haugen, Robert A, and Nardin L Baker, 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401–439.

Hawkins, Eugene H, Stanley C Chamberlin, and Wayne E Daniel, 1984, Earnings expectations and security prices, *Financial Analysts Journal* 40, 24–38.

Heston, Steven L, and Ronnie Sadka, 2008, Seasonality in the cross-section of stock returns, *Journal of Financial Economics* 87, 418–445.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets?, *Journal of Accounting and Economics* 38, 297–331.

Horel, Enguerrand, and Kay Giesecke, 2019, Towards explainable ai: Significance tests for neural networks, *arXiv preprint arXiv:1902.06021* .

Hou, Kewei, G Andrew Karolyi, and Bong-Chan Kho, 2011, What factors drive global stock returns?, *The Review of Financial Studies* 24, 2527–2574.

Hou, Kewei, and David T Robinson, 2006, Industry concentration and average stock returns, *The Journal of Finance* 61, 1927–1956.

Hou, Kewei, Chen Xue, and Lu Zhang, 2017, Replicating anomalies, Technical report, National Bureau of Economic Research.

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen, 1995, Market underreaction to open market share repurchases, *Journal of financial economics* 39, 181–208.

Ince, Ozgur S, and R Burt Porter, 2006, Individual equity return data from thomson datastream: Handle with care!, *Journal of Financial Research* 29, 463–479.

Jacobs, Heiko, and Sebastian Müller, 2017a, ... and nothing else matters? on the dimensionality and predictability of international stock returns .

Jacobs, Heiko, and Sebastian Müller, 2017b, Anomalies across the globe: Once public, no longer existent? .

Jacobs, Heiko, and Sebastian Müller, 2017c, Measuring mispricing .

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *The Journal of finance* 45, 881–898.

Jegadeesh, Narasimhan, Joonghyuk Kim, Susan D Krische, and Charles Lee, 2004, Analyzing the analysts: When do recommendations add value?, *The journal of finance* 59, 1083–1124.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of finance* 48, 65–91.

Kelly, Bryan, and Hao Jiang, 2014, Tail risk and asset prices, *The Review of Financial Studies* 27, 2841–2871.

Kot, Hung Wan, and Kalok Chan, 2006, Can contrarian strategies improve momentum profits .

Kyle, Albert S, and Anna A Obizhaeva, 2016, Market microstructure invariance: Empirical hypotheses, *Econometrica* 84, 1345–1404.

La Porta, Rafael, 1996, Expectations and the cross-section of stock returns, *The Journal of Finance* 51, 1715–1742.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, 1994, Contrarian investment, extrapolation, and risk, *The journal of finance* 49, 1541–1578.

Lam, FY Eric C, and KC John Wei, 2011, Limits-to-arbitrage, investment frictions, and the asset growth anomaly, *Journal of Financial Economics* 102, 127–149.

Lee, Charles, and Bhaskaran Swaminathan, 2000, Price momentum and trading volume, *the Journal of Finance* 55, 2017–2069.

Lee, Kuan-Hui, 2011, The world price of liquidity risk, *Journal of Financial Economics* 99, 136–161.

Lewellen, Jonathan, et al., 2015, The cross-section of expected stock returns, *Critical Finance Review* 4, 1–44.

Li, Dongmei, 2011, Financial constraints, r&d investment, and stock returns, *The Review of Financial Studies* 24, 2974–3007.

Lockwood, Larry, and Wikrom Prombutr, 2010, Sustainable growth and stock returns, *Journal of Financial Research* 33, 519–538.

Loughran, Tim, and Jay W Wellman, 2011, New evidence on the relation between the enterprise multiple and average stock returns, *Journal of Financial and Quantitative Analysis* 46, 1629–1650.

Lyandres, Evgeny, Le Sun, and Lu Zhang, 2007, The new issues puzzle: Testing the investment-based explanation, *The Review of Financial Studies* 21, 2825–2855.

McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.

McLean, R David, Jeffrey Pontiff, and Akiko Watanabe, 2009, Share issuance and cross-sectional returns: International evidence, *Journal of Financial Economics* 94, 1–17.

Moskowitz, Tobias J, and Mark Grinblatt, 1999, Do industries explain momentum?, *The Journal of Finance* 54, 1249–1290.

Novy-Marx, Robert, 2010, Operating leverage, *Review of Finance* 15, 103–134.

Novy-Marx, Robert, 2012, Is momentum really momentum?, *Journal of Financial Economics* 103, 429–453.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Novy-Marx, Robert, and Mihail Velikov, 2015, A taxonomy of anomalies and their trading costs, *The Review of Financial Studies* 29, 104–147.

Ortiz-Molina, Hernán, and Gordon M Phillips, 2014, Real asset illiquidity and the cost of capital, *Journal of Financial and Quantitative Analysis* 49, 1–32.

Palazzo, Berardino, 2012, Cash holdings, risk, and expected returns, *Journal of Financial Economics* 104, 162–185.

Penman, Stephen H, Scott A Richardson, and Irem Tuna, 2007, The book-to-price effect in stock returns: accounting for leverage, *Journal of Accounting Research* 45, 427–467.

Piotroski, Joseph D, 2000, Value investing: The use of historical financial statement information to separate winners from losers, *Journal of Accounting Research* 1–41.

Pontiff, Jeffrey, and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, *The Journal of Finance* 63, 921–945.

Richardson, Scott A, Richard G Sloan, Mark T Soliman, and Irem Tuna, 2006, The implications of accounting distortions and growth for accruals and profitability, *The Accounting Review* 81, 713–743.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *The Journal of finance* 39, 1127–1139.

Rouwenhorst, K Geert, 1998, International momentum strategies, *The journal of finance* 53, 267–284.

Rouwenhorst, K Geert, 1999, Local return factors and turnover in emerging stock markets, *The journal of finance* 54, 1439–1464.

Sirignano, Justin, Apaar Sadhwani, and Kay Giesecke, 2016, Deep learning for mortgage risk, *arXiv preprint arXiv:1607.02470* .

Sloan, Allan, 1996, Create an account or log in, *Accounting review* 71, 289–315.

Soliman, Mark T, 2008, The use of dupont analysis by market participants, *The Accounting Review* 83, 823–853.

Spiess, D Katherine, and John Affleck-Graves, 1995, Underperformance in long-run stock returns following seasoned equity offerings, *Journal of Financial Economics* 38, 243–267.

Thomas, Jacob K, and Huai Zhang, 2002, Inventory changes and future returns, *Review of Accounting Studies* 7, 163–187.

Tibshirani, Robert, 1996, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)* 267–288.

Titman, Sheridan, KC John Wei, and Feixue Xie, 2004, Capital investments and stock returns, *Journal of financial and Quantitative Analysis* 39, 677–700.

Titman, Sheridan, KC John Wei, and Feixue Xie, 2013, Market development and the asset growth effect: International evidence, *Journal of Financial and Quantitative Analysis* 48, 1405–1432.

Tobek, Ondrej, and Martin Hronec, 2018, Does source of fundamental data matter? .

Tuzel, Selale, 2010, Corporate real estate holdings and the cross-section of stock returns, *The Review of Financial Studies* 23, 2268–2302.

Watanabe, Akiko, Yan Xu, Tong Yao, and Tong Yu, 2013, The asset growth effect: Insights from international equity markets, *Journal of Financial Economics* 108, 529–563.

Whited, Toni M, and Guojun Wu, 2006, Financial constraints risk, *The Review of Financial Studies* 19, 531–559.

Zhang, X, 2006, Information uncertainty and stock returns, *The Journal of Finance* 61, 105–137.

Online Appendix for Does It Pay to Follow Anomalies Research? Machine Learning Approach with International Evidence

July 2018

Table IA1
Industries in Datastream Level 3 Classification and Corresponding Four Digit SIC

Datastream lvl 3 industry	SIC codes
Automobiles & Parts	3011, 3510, 3714, 3751, 5013
Basic Resources	800, 1000, 1040, 1090, 1220, 1221, 2421, 2600, 2611, 2621, 2631, 3310, 3312, 3317, 3330, 3334, 3350, 3360, 3444, 3460, 3720, 5050, 5051
Chemicals	2810, 2820, 2821, 2833, 2851, 2860, 2870, 2890, 2891, 2990, 3080, 3081, 3341, 5160
Construct. & Material	1400, 1540, 1600, 1623, 1731, 2400, 2430, 2950, 3211, 3231, 3241, 3250, 3270, 3272, 3281, 3290, 3430, 3440, 3442, 3448, 5031, 5070, 5072
Financial Services(3)	6111, 6141, 6153, 6159, 6162, 6163, 6172, 6189, 6200, 6211, 6221, 6282, 6361, 6500, 6510, 6770, 6795, 6798, 6799, 8880, 8888, 9995
Food & Beverage	100, 200, 900, 2000, 2011, 2013, 2015, 2020, 2024, 2030, 2033, 2040, 2050, 2052, 2060, 2070, 2080, 2082, 2086, 2090, 2092
Healthcare	2590, 2800, 2834, 2835, 2836, 3060, 3821, 3826, 3841, 3842, 3843, 3844, 3845, 3851, 4100, 5047, 6324, 8000, 8011, 8050, 8051, 8060, 8062, 8071, 8082, 8090, 8093, 8300, 8731
Ind. Goods & Services	1700, 2390, 2650, 2670, 2673, 2750, 2761, 3050, 3086, 3089, 3221, 3320, 3357, 3390, 3411, 3412, 3443, 3451, 3452, 3470, 3480, 3490, 3523, 3524, 3530, 3531, 3532, 3537, 3540, 3541, 3550, 3555, 3560, 3561, 3562, 3564, 3567, 3569, 3575, 3580, 3585, 3590, 3600, 3612, 3613, 3620, 3621, 3634, 3640, 3669, 3670, 3672, 3677, 3678, 3679, 3690, 3711, 3713, 3715, 3721, 3724, 3728, 3730, 3743, 3760, 3812, 3822, 3823, 3824, 3825, 3827, 3829, 3861, 3910, 4011, 4013, 4210, 4213, 4231, 4400, 4412, 4513, 4700, 4731, 4950, 4953, 4955, 4961, 5000, 5063, 5065, 5080, 5082, 5084, 5090, 5099, 6099, 6794, 7320, 7350, 7359, 7361, 7363, 7374, 7377, 7380, 7381, 7384, 7385, 7389, 7829, 8111, 8200, 8351, 8600, 8700, 8711, 8734, 8741, 8742, 8744, 9721
Insurance	6311, 6321, 6331, 6351, 6411
Media	2711, 2721, 2731, 2732, 2741, 2780, 4832, 4833, 4841, 7310, 7311, 7330, 7331, 7819, 7822, 8900
Oil & Gas	1311, 1381, 1382, 1389, 2911, 3533, 4522, 4610, 4900, 5171, 5172, 6792
Pers & Househld Goods	1531, 2100, 2111, 2200, 2211, 2221, 2250, 2253, 2273, 2300, 2320, 2330, 2340, 2451, 2452, 2510, 2511, 2520, 2522, 2531, 2540, 2771, 2840, 2842, 2844, 3021, 3100, 3220, 3260, 3420, 3433, 3630, 3651, 3716, 3790, 3873, 3911, 3931, 3942, 3944, 3949, 3950, 3960, 5020, 5030, 5064, 5130, 5150, 5190, 6552
Real Estate	6519, 6531
Retail	700, 2790, 3140, 4220, 5094, 5010, 5110, 5122, 5140, 5141, 5180, 5200, 5211, 5271, 5311, 5331, 5399, 5400, 5411, 5412, 5500, 5531, 5600, 5621, 5651, 5661, 5700, 5712, 5731, 5734, 5735, 5912, 5940, 5944, 5945, 5960, 5961, 5990, 6399, 7200, 7340, 7500, 7600, 7841
Technology	3559, 3570, 3571, 3572, 3576, 3577, 3578, 3579, 3661, 3663, 3674, 3695, 4899, 5040, 5045, 7370, 7371, 7372, 7373
Telecommunications	4812, 4813, 4822
Travel & Leisure	1520, 3652, 3990, 4512, 4581, 5810, 5812, 6512, 6513, 6532, 7000, 7011, 7510, 7812, 7830, 7900, 7948, 7990, 7997
Utilities	4911, 4922, 4923, 4924, 4931, 4932, 4941, 4991, 5900
Banks	6021, 6022, 6029, 6035, 6036, 6199

Table IA2
Performance of the Mispricing Strategy Estimated on the Liquid Sample

The table shows returns of the mispricing strategy described in Table IV that is estimated on the individual stocks from the US. The results labelled "Estimated on All-but-micro-caps" are taken from Table IV where the predictive regressions were estimated on all-but-micro-caps sample of stocks while the predictive regressions for results labelled "Estimated on Liquid Universe of Stock" were estimated on a more liquid universe of stocks. See Section I.A for definition of the two samples. Panel A describes results from weighted least squares estimation method, Panel B from gradient boosting regression trees method, and Panel C from neural networks. The returns are reported in percentage points per month.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Panel A: Weighted Least Squares										
Estimated on All-but-micro-caps										
Mean	0.801	0.680	0.922	0.782	0.810	0.575	0.647	0.648	0.633	0.639
Sharpe Ratio	0.479	0.541	0.701	0.497	0.763	0.348	0.472	0.410	0.318	0.550
Skewness	-0.340	0.239	-0.425	-0.356	0.025	-0.121	-0.017	-0.688	-0.238	-0.041
Kurtosis	8.521	5.673	4.167	3.713	8.488	7.214	6.483	5.544	4.824	7.880
Max Drawdown	-64.70	-37.10	-43.36	-47.61	-43.51	-69.75	-34.16	-44.52	-49.86	-50.85
Estimated on Liquid Universe of Stocks										
Mean	0.751	0.751	0.909	0.854	0.797	0.776	0.745	0.903	0.798	0.844
Sharpe Ratio	0.435	0.572	0.706	0.500	0.732	0.469	0.577	0.595	0.424	0.747
Skewness	-0.459	-0.190	-0.157	-0.600	-0.359	0.123	-0.237	-0.005	-0.968	0.042
Kurtosis	7.245	6.560	4.840	6.997	7.957	6.983	5.575	3.734	10.38	6.720
Max Drawdown	-48.76	-47.03	-33.82	-40.37	-32.49	-56.85	-30.22	-39.31	-47.27	-30.62
Information Ratio	-0.045	0.064	-0.013	0.053	-0.017	0.153	0.082	0.206	0.101	0.219
Panel B: Gradient Boosting Regression Trees										
Estimated on All-but-micro-caps										
Mean	1.165	0.870	1.173	1.650	1.163	1.391	0.591	1.011	1.415	1.033
Sharpe Ratio	0.720	0.644	0.766	1.005	1.146	0.831	0.412	0.525	0.800	0.870
Skewness	0.319	-1.160	0.682	-0.437	-0.449	0.561	-1.314	0.800	-0.112	-0.433
Kurtosis	6.653	10.17	8.274	5.575	6.812	9.287	12.03	8.611	4.718	7.797
Max Drawdown	-38.31	-48.25	-34.37	-36.65	-27.45	-43.93	-42.31	-41.79	-39.58	-35.62
Estimated on Liquid Universe of Stocks										
Mean	1.110	0.808	0.907	1.179	1.010	1.219	0.979	0.906	1.295	1.111
Sharpe Ratio	0.757	0.692	0.835	0.719	1.103	0.786	0.745	0.564	0.566	1.047
Skewness	-0.548	-0.235	-0.822	0.731	-0.720	-0.116	0.630	1.380	2.613	0.001
Kurtosis	9.152	7.028	6.618	8.156	9.068	5.293	7.189	19.79	31.79	5.125
Max Drawdown	-34.33	-37.94	-25.85	-37.13	-23.71	-33.44	-26.73	-43.35	-52.74	-24.79
Information Ratio	-0.042	-0.066	-0.207	-0.289	-0.217	-0.111	0.320	-0.066	-0.052	0.087
Panel C: Neural Networks										
Estimated on All-but-micro-caps										
Mean	1.416	1.097	1.295	1.752	1.346	1.420	0.826	1.100	1.177	1.093
Sharpe Ratio	0.905	0.880	1.130	1.086	1.582	0.905	0.649	0.693	0.697	1.042
Skewness	-0.083	-0.082	-0.149	0.244	-0.310	-0.167	-0.470	0.629	0.638	-0.255
Kurtosis	7.316	4.827	4.446	5.091	5.304	6.432	7.050	10.37	5.075	6.806
Max Drawdown	-44.60	-33.93	-24.70	-38.10	-18.90	-48.11	-31.93	-37.09	-54.45	-33.25
Estimated on Liquid Universe of Stocks										
Mean	1.248	0.835	1.001	1.196	1.103	1.454	0.670	0.810	1.095	1.102
Sharpe Ratio	0.674	0.617	0.812	0.700	0.939	0.843	0.491	0.519	0.557	0.981
Skewness	-0.069	0.032	-0.506	-0.306	-0.150	0.174	-0.272	-0.612	-0.159	-0.283
Kurtosis	6.244	5.382	4.654	4.115	6.338	5.088	6.069	4.077	8.655	5.986
Max Drawdown	-51.51	-41.31	-32.50	-56.63	-33.54	-35.68	-38.76	-52.72	-66.17	-27.51
Information Ratio	-0.083	-0.209	-0.261	-0.345	-0.224	0.017	-0.124	-0.185	-0.044	0.007

Table IA3
Performance of the Mispricing Strategy Estimated on Stocks Outside the US: Weighted Least Squares Regressions

The table shows returns of the mispricing strategy as described in Table IV that is estimated on individual stocks from the US, US & Japan, US & Europe, or US & Japan & Europe. The returns are in percentage points per month.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Evidence from the US										
Mean	0.801	0.680	0.922	0.782	0.810	0.575	0.647	0.648	0.633	0.639
Sharpe Ratio	0.479	0.541	0.701	0.497	0.763	0.348	0.472	0.410	0.318	0.550
Skewness	-0.340	0.239	-0.425	-0.356	0.025	-0.121	-0.017	-0.688	-0.238	-0.041
Kurtosis	8.521	5.673	4.167	3.713	8.488	7.214	6.483	5.544	4.824	7.880
Max Drawdown	-64.70	-37.10	-43.36	-47.61	-43.51	-69.75	-34.16	-44.52	-49.86	-50.85
Evidence from the US & Japan										
Mean	0.694	0.648	0.796	0.748	0.722	0.607	0.703	0.602	0.569	0.639
Sharpe Ratio	0.404	0.485	0.543	0.450	0.642	0.337	0.485	0.339	0.302	0.498
Skewness	-0.215	0.296	-0.407	-0.373	0.055	-0.167	0.296	0.590	0.105	0.362
Kurtosis	7.929	5.619	5.593	4.571	7.920	7.716	5.202	11.16	3.816	9.129
Max Drawdown	-69.18	-42.84	-55.39	-48.93	-49.92	-73.56	-36.78	-53.55	-51.64	-54.34
Information Ratio	-0.181	-0.056	-0.189	-0.035	-0.218	0.042	0.098	-0.044	-0.045	0.000
Evidence from the US & Europe										
Mean	0.780	0.700	0.854	1.140	0.842	0.708	0.638	0.569	0.897	0.715
Sharpe Ratio	0.435	0.546	0.646	0.697	0.736	0.418	0.477	0.340	0.487	0.603
Skewness	-0.581	0.082	-0.590	-0.356	-0.397	-0.308	0.141	-0.617	-0.255	-0.221
Kurtosis	10.31	5.844	4.491	3.826	8.745	8.525	5.838	4.677	3.726	7.486
Max Drawdown	-63.84	-41.66	-41.74	-48.17	-41.68	-63.39	-36.58	-58.68	-49.51	-44.83
Information Ratio	-0.037	0.031	-0.116	0.442	0.082	0.173	-0.014	-0.098	0.274	0.142
Evidence from the US & Japan & Europe										
Mean	0.808	0.699	0.884	1.112	0.853	0.765	0.626	0.469	1.020	0.730
Sharpe Ratio	0.444	0.516	0.645	0.691	0.736	0.422	0.426	0.267	0.530	0.580
Skewness	-0.359	0.304	-0.267	-0.321	-0.096	-0.303	-0.038	0.134	-0.466	0.034
Kurtosis	8.281	5.359	4.212	4.241	7.407	7.961	5.712	7.329	5.093	7.344
Max Drawdown	-70.99	-45.05	-39.66	-42.05	-45.60	-72.31	-37.82	-63.48	-45.19	-50.12
Information Ratio	0.009	0.028	-0.057	0.346	0.096	0.221	-0.030	-0.167	0.320	0.155

Table IA4
Is Marginal Return to Following New Anomalies Decreasing over Time?

The table shows returns of the mispricing strategy described in Table IV that is estimated on the individual stocks from the US. The set of anomalies in the estimation is restricted to those that were published before 1995, 2000, 2005, or 2010. Returns are reported in percentage points per month over the 2010-2016 period.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Weighted Least Squares										
Published by 1995										
Mean	0.189	0.408	0.324	0.343	0.301	0.190	0.300	1.028	0.424	0.337
Sharpe Ratio	0.210	0.408	0.306	0.255	0.440	0.208	0.263	0.926	0.295	0.442
Published by 2000										
Mean	0.193	0.777	0.577	0.055	0.409	0.129	0.588	1.063	-0.121	0.415
Sharpe Ratio	0.213	0.811	0.475	0.042	0.592	0.138	0.506	0.816	-0.090	0.540
Information Ratio	0.006	0.556	0.333	-0.306	0.253	-0.086	0.354	0.040	-0.421	0.150
Published by 2005										
Mean	0.896	0.780	0.873	0.517	0.793	0.807	0.476	1.376	0.010	0.860
Sharpe Ratio	1.031	0.812	0.760	0.386	1.181	1.001	0.392	1.091	0.006	1.189
Information Ratio	1.233	0.004	0.476	0.380	0.974	0.858	-0.107	0.376	0.081	0.686
Published by 2010										
Mean	0.997	0.868	0.978	1.283	0.994	0.797	0.569	0.934	1.247	0.892
Sharpe Ratio	1.240	0.997	0.978	0.937	1.589	0.907	0.561	0.830	0.863	1.262
Information Ratio	0.198	0.156	0.153	0.796	0.595	-0.015	0.130	-0.539	0.792	0.069
Gradient Boosting Regression Trees										
Published by 1995										
Mean	0.514	0.446	0.634	1.034	0.609	0.446	0.695	0.706	0.817	0.617
Sharpe Ratio	0.638	0.480	0.637	0.723	1.115	0.483	0.739	0.661	0.528	1.006
Published by 2000										
Mean	0.517	0.906	1.090	1.595	0.918	0.134	0.652	1.172	1.399	0.561
Sharpe Ratio	0.655	1.015	1.167	1.246	1.765	0.150	0.728	1.001	1.049	0.947
Information Ratio	0.006	0.658	0.487	0.546	0.752	-0.389	-0.054	0.348	0.432	-0.096
Published by 2005										
Mean	0.830	0.948	1.169	1.583	1.045	0.822	0.792	1.182	1.161	0.941
Sharpe Ratio	1.223	1.080	1.286	1.483	2.324	1.054	1.082	1.241	1.028	1.978
Information Ratio	0.483	0.057	0.095	-0.009	0.294	0.827	0.154	0.009	-0.152	0.645
Published by 2010										
Mean	1.011	1.140	0.817	1.943	1.121	0.509	1.085	0.911	1.765	0.898
Sharpe Ratio	1.443	1.264	0.803	1.562	2.039	0.585	1.221	0.733	1.435	1.530
Information Ratio	0.348	0.289	-0.447	0.369	0.216	-0.405	0.382	-0.277	0.485	-0.094

Appendix A. Monthly Updated Fundamental Anomalies

Anomalies based on annual financial statements have so far been updated annually every June. June was chosen so that firms with financial year ending in December have 6 months to publish their statements. The explicit assumption was that all the firms publish their statements within 6 months after their financial year has ended. The rule was originally devised on the US data where great majority of firms have their financial year ending in December. The usual financial year end is, however, different in the other regions. 78% of firms in Japan have financial year ending in March. The most frequent choice of financial year end in Asia Pacific region is either December or June both being about equally likely. Financial year end date outside December leads to the financial statements being older than 6 months in June and thus being less relevant. Bartram and Grinblatt (2018) and Jacobs and Müller (2017c) circumvented this problem when working with international data by relying on point-in-time Reuters database that presents financial statements as they were published by a given date and creating the fundamental signals monthly. We do not have access to the point-in-time database but we will here create a pseudo point-in-time database and will also refresh the fundamental signals monthly rather than annually.

Table IA5 presents results from Table IV based on the annual construction of fundamental signals along with their monthly construction. Everything remains the same as in Table IV with the only difference being that the fundamental signals are updated every month with financial statement information from financial years ending at least 6 months prior. The explicit assumption again is that all the firms publish their statements within the 6 months after their financial year has ended. All the trade data information such as market cap is also updated monthly and taken the most recent. Market cap was previously taken from the previous calendar year end as in Fama and French (1992) and was therefore outdated by 6 months by June. Asness and Frazzini (2013) showed that market cap from June leads to better performance of value factor. There can therefore also be some benefit from shifting the trade data information.

Table IA5 documents that the lag in availability of the financial statements leads to some loss in performance in almost all the regions. Both mean returns and Sharpe ratios with the monthly updating of the fundamental signals are about 10% higher relative to when they are updated annually. To conclude, the monthly updating can slightly improve the performance of the mispricing strategy but it does not affect the main conclusions of this study.

Table IA5
Performance of the Mispricing Strategy with Monthly Updated Fundamental Signals

The table shows returns of the mispricing strategy described in Table IV that is estimated on the individual stocks from the US. The results labelled "Annually Updated Fundamental Signals" directly correspond to Table IV where the fundamental signals are updated every June while the results labelled "Monthly Updated Fundamental Signals" are created using fundamental signals that are updated every month based on financial statements released more than six months prior. Panel A describes results from weighted least squares estimation method while Panel B reports results from gradient boosting regression trees method. The returns are reported in percentage points per month.

	Equal-weighted					Value-weighted				
	USA	Europe	Japan	AP	Global	USA	Europe	Japan	AP	Global
Panel A: Weighted Least Squares										
Annually Updated Fundamental Signals										
Mean	0.801	0.680	0.922	0.782	0.810	0.575	0.647	0.648	0.633	0.639
Sharpe Ratio	0.479	0.541	0.701	0.497	0.763	0.348	0.472	0.410	0.318	0.550
Skewness	-0.340	0.239	-0.425	-0.356	0.025	-0.121	-0.017	-0.688	-0.238	-0.041
Kurtosis	8.521	5.673	4.167	3.713	8.488	7.214	6.483	5.544	4.824	7.880
Max Drawdown	-64.70	-37.10	-43.36	-47.61	-43.51	-69.75	-34.16	-44.52	-49.86	-50.85
Monthly Updated Fundamental Signals										
Mean	0.889	0.750	0.869	1.078	0.883	0.736	0.585	0.640	0.676	0.696
Sharpe Ratio	0.537	0.644	0.690	0.663	0.867	0.447	0.440	0.386	0.360	0.634
Skewness	-0.203	0.246	-0.479	-0.307	0.008	-0.046	0.075	-0.139	-0.209	-0.063
Kurtosis	8.039	4.970	4.987	3.822	7.683	6.466	6.056	5.155	4.145	6.239
Max Drawdown	-63.40	-38.59	-39.29	-42.31	-37.78	-65.32	-26.37	-62.52	-41.22	-41.57
Information Ratio	0.173	0.133	-0.074	0.272	0.219	0.218	-0.105	-0.008	0.029	0.114
Panel B: Gradient Boosting Regression Trees										
Annually Updated Fundamental Signals										
Mean	1.165	0.870	1.173	1.650	1.163	1.391	0.591	1.011	1.415	1.033
Sharpe Ratio	0.720	0.644	0.766	1.005	1.146	0.831	0.412	0.525	0.800	0.870
Skewness	0.319	-1.160	0.682	-0.437	-0.449	0.561	-1.314	0.800	-0.112	-0.433
Kurtosis	6.653	10.17	8.274	5.575	6.812	9.287	12.03	8.611	4.718	7.797
Max Drawdown	-38.31	-48.25	-34.37	-36.65	-27.45	-43.93	-42.31	-41.79	-39.58	-35.62
Monthly Updated Fundamental Signals										
Mean	1.264	1.039	1.242	1.597	1.254	1.492	0.771	1.207	1.314	1.125
Sharpe Ratio	0.786	0.840	0.858	0.965	1.241	0.900	0.538	0.645	0.770	0.902
Skewness	0.260	-0.848	0.438	-0.099	-0.337	1.041	-1.332	1.057	0.203	0.518
Kurtosis	7.663	10.95	8.076	5.330	6.896	8.525	16.78	10.60	5.307	7.435
Max Drawdown	-47.13	-43.82	-30.79	-37.49	-27.12	-43.11	-43.00	-36.56	-35.74	-31.02
Information Ratio	0.142	0.270	0.091	-0.042	0.237	0.110	0.179	0.191	-0.065	0.141